Human tumor-associated macrophage and monocyte transcriptional landscapes reveal cancer-specific reprogramming, biomarkers and therapeutic targets

Human tumor-associated macrophage and monocyte transcriptional landscapes reveal cancer-specific reprogramming, biomarkers and therapeutic targets

Luca Cassetta[1, 12], Stamatina Fragkogianni[1, 12], Andrew H. Sims[2], Agnieszka Swierczak[1],Lesley M. Forrester[4], Hui Zhang[5], Daniel Soong[1], Tiziana Cotechini[3], Pavana Anur[4], Elaine Lin[3], Antonella Fidanza[4], Martha Lopez-Yrigoyen[4], Michael R. Millar[1,9], Alexandra Urman[6], , Zhichao Ai[1], Paul T. Spellman[10], E. Shelley Hwang[11], J. Michael Dixon[6], Lisa Wiechmann[7], Lisa M. Coussens[3], Harriet W. Smith[8] and Jeffrey W. Pollard[1,5,*]

[1]MRC Centre for Reproductive Health, Queen's Medical Research Institute, The University of Edinburgh, Edinburgh, UK

[2]Applied Bioinformatics of Cancer, University of Edinburgh Cancer Research Centre, Institute of Genetics and Molecular Medicine, Edinburgh, UK

[3]Department of Cell, Developmental & Cancer Biology, and Knight Cancer Institute, Oregon Health & Science University, Portland, USA

[4]MRC Centre for Regenerative Medicine, University of Edinburgh, Edinburgh, UK

[5]Department of Developmental and Molecular Biology, Albert Einstein College of Medicine, New York, USA

[6]Edinburgh Breast Unit and Breast Cancer Now Research Unit, University of Edinburgh, Edinburgh, UK

[7]Department of Surgery, Montefiore Medical College, New York, USA

[8]Department of Obstetrics and Gynecology, Albert Einstein College of Medicine and Montefiore Medical Center, New York, USA

[9]Aquila Biomedical, Edinburgh Bioquarter, Little France Road, Edinburgh, UK

[10]Department of Molecular and Medical Genetics and Knight Cancer Institute, Oregon Health & Science University, Portland, USA

[11]Department of Surgery, Duke University Medical Center, Durham, North Carolina, USA

[12]Co-first authors

[*]Correspondence: Jeff.Pollard@ed.ac.uk.

Summary

The roles of tumor-associated macrophages (TAMs) and circulating monocytes in human cancer are poorly understood. Here, we show that monocyte subpopulation distribution and transcriptomes are significantly altered by the presence of endometrial and breast cancer. Furthermore, TAMs from endometrial and breast cancers are transcriptionally distinct from monocytes and their respective tissue resident macrophages. We identified a breast TAM signature that is highly enriched in aggressive breast cancer subtypes and associated with shorter disease-specific survival. We also identified an auto-regulatory loop between TAMs and cancer cells driven by TNFα involving *SIGLEC1* and *CCL8,* that is self-reinforcing through the production of CSF1. Together these data provide direct evidence that monocyte and macrophage transcriptional landscapes are perturbed by cancer, reflecting patient outcomes.

Introduction

Tumours evolve as ecosystems consisting of tumor, stromal and infiltrating immune cells. Macrophages are major components of this ecosystem. In mouse models of cancer, different subpopulations of TAMs promote angiogenesis, tumour cell invasion, intravasation, and at the metastatic site, tumour cell extravasation and persistent growth and suppress anti-tumor cytolytic T cell responses. TAMs can also promote chemo-, radio-therapy and immuno-therapy resistance (Cassetta and Pollard, 2018). In homeostasis, tissue macrophages have different origins including yolk sac, fetal liver and bone marrow. However in most, but not all, mouse cancer models, TAMs are recruited from bone marrow progenitors known as monocytes. These monocytes are termed classical (human CD14++CD16- and mouse CD11b+Ly6C+) and non-classical (human CD14+CD16+; mouse CD11b+Ly6C). The classical population is recruited as the tumour progresses and differentiates in situ to TAMs often via a CCL2-CCR2 chemokine signalling pathway. Inhibition of CCR2 signaling blocks TAM recruitment and thus inhibits tumor cell seeding and persistent growth improving the survival of mice (Qian et al., 2011). Transcriptional profiling of total isolated monocytes from renal carcinoma and colorectal cancer patients showed distinct proinflammatory and pro-tumoral profiles compared to healthy individuals (Chittezhath et al., 2014; Hamm et al., 2015). In breast cancer, higher numbers of non-classical are negatively associated with tumour size and disease stage (Feng et al., 2011). However, in depth characterization of human monocytes in breast and endometrial cancer is lacking.

The pro-tumoural behavior of monocytes and TAMs in mouse models has made them attractive therapeutic targets. Targeting strategies include inhibiting monocyte recruitment, depletion of TAMs and functional/phenotypic reprogramming (Cassetta and Pollard, 2018). These therapies however are limited by the lack of TAM-specific markers (Williams et al., 2016) as well as our limited understanding of their functions in human cancers (Takeya and Komohara, 2016). To address these issues, we analysed the transcriptional profiles of circulating monocytes and TAMs in breast and endometrial cancer. We observed profound changes in the transcriptomes of cancer monocytes and TAMs compared to their normal healthy counterparts. Circulating monocytes were profoundly different from their progeny in tissues showing the influence of the local tissue microenvironment. This analysis allowed identification of a TAM signature associated with poor clinical outcomes and two breast cancer TAM markers, CCL8 and SIGLEC1, which have prognostic value and represent potential therapeutic targets.

Results

Cancer alters the transcriptome of human monocytes

We isolated total monocytes from women with breast (N = 32), endometrial cancer (N = 3) and healthy controls (N = 45) and performed bulk RNA-sequencing (RNA-seq) (**Figure S1A-B**). Although there are outliers, principal components analysis (PCA) and hierarchical clustering segregated the transcriptomic profiles of normal monocytes (Mo) from that of monocytes from breast or endometrial cancer patients (**Figure 1 A, B**). Thus we designated cancer monocytes as Tumor-Educated Monocytes (TEMo). Limma differential expression analysis (DEA) revealed 865 differentially expressed genes (DEGs) in breast TEMo compared to Mo (543 up and 322 down; FDR <= 0.05, **File S1B**) and 997 DEGs in endometrial TEMo compared to Mo (498 up and 499 down; FDR <= 0.05, **File S1B**). Due to the limited size of endometrial TEMo samples, we focused our downstream analysis on the breast TEMo population. Gene ontology analysis reported a number of enriched GO terms such as cell migration, angiogenesis, cell communication and apoptotic process (**Figure 1C**). A number of genes encoding transmembrane receptors, soluble factors, transcription factors and enzymes were deregulated, including increased expression of transcripts encoding immune regulatory receptors (*CD200R1*), pro-apoptotic molecules (*TNFSF10*, *TRAIL*) and pro-angiogenic factors (Hepatocyte Growth Factor, *HGF*; Angiopoietin1, *ANGPT1*) (**Figure 1D**). Quantitative RT-PCR (qPCR) of monocytic RNA derived from an independent breast cancer cohort confirmed significant increased expression of these genes as independent validation of the RNA-seq analysis (**Figure 1E**).

To better understand if this shift in the transcriptomes of monocytes during malignancy was driven by a specific subpopulation, we collected and analysed by flow cytometry classical and non-classical monocytes from two independent cancer cohorts (Cohort 1: endometrial and breast, Bronx, NY, USA; Cohort 2: breast, Edinburgh, UK) as well as healthy women (**Figure 1F and Figure S1C-E, see File S2 for clinical information**). Non-classical monocytes from cancer patients exhibited a significant expansion compared to healthy controls in both cohorts without significant differences between endometrial and breast cancer patients (**Figure 1F**). This expansion was associated with a significant increase in CX3CL1 and reduction of CCL2 in cancer patients' sera (**Figure 1G**). The expression level of the main receptor of CCL2, CCR2, did not change among subpopulations and conditions although CX3CR1, the CX3CL1 receptor, was significantly downregulated in classical monocytes from cancer patients compared to control consistent with the alterations in monocytic populations (**Figure 1H**). Next, we isolated non-classical monocytes from 13 cancer patients (N = 6 breast and N = 7 endometrial) and 5 healthy women and performed RNA-seq. PCA and hierarchical clustering revealed distinct non-classical monocyte clusters in cancer patients versus healthy volunteers (**Figure S1F**). Limma DEA revealed 139 DEGs in non-classical monocytes from breast cancer patients

compared to healthy individuals (103 up and 36 down; FDR <= 0.05, **File S1C**). Similarly, we identified 576 DEGs in non-classical monocytes in endometrial cancer patients compared to healthy individuals (501 up and 75 down; FDR <= 0.05, **File S1D**). Hierarchical clustering showed similar patterns of gene expression changes in non-classical monocytes from women with breast and endometrial cancer compared to healthy women (**Figure S1G**).

Given the significant transcriptional differences in monocytes between cancer patients and healthy volunteers, we hypothesised that a TEMo signature from a liquid biopsy with minimal processing could be generated for breast cancer detection. We tested this hypothesis using total monocytes and a Recursive Feature Elimination with a Random Forest (RFE-RF) algorithm. The dataset was split into training (70%, N = 55, 32 healthy, 23 cancer samples) and testing (30%, N = 22, 13 healthy, 9 cancer samples) sets; the training set was used for feature selection and model training using 5 times 10-fold cross validation (CV). In the training set, the algorithm selected 17 highest performing genes that yielded an average of 85% accuracy, 88% sensitivity and 83% specificity during CV (**Figure S1H**). Subsequent validation using the holdout test set yielded 82% accuracy, 100% sensitivity and 69% specificity (**Figure 1I-K**) and area under curve (AUC) of 96% to detect cancer (**Figure 1L**). In contrast, random classifiers, as determined by 1000 rounds of random class permutations during Random Forest model training, had no predictive power (mean accuracy: 53%, SD ± 6.8%, p = 0.001) and significantly different from to the 17-gene predictive signature above (**Figure S1I**).

Gene expression profiles of TAMs in human breast and endometrial cancers

There is significant evidence showing pro-tumoral profiles of TAMs in mouse models of cancer; however, a detailed characterization of their transcriptomes and phenotypes in human cancers is still lacking. Thus, we analysed TAM transcriptomes by RNA-seq from breast and endometrial cancer in comparison to resident macrophages from homeostatic tissue after FACS sorting (**Figure S2A**). PCA and hierarchical clustering revealed distinct clusters of breast tissue resident macrophages (Br-RM) and breast cancer TAMs (Br-TAM) (**Figure 2A, B**). Limma DEA revealed 1873 DEGs in Br-TAM compared with Br-RM (1301 up and 572 down; FDR <= 0.05, **File S1E**). Gene ontology analysis reported several enriched GO terms such as cell motility and activation, vasculature development and immune response (**Figure 2C**). Br-TAM showed increased transcript abundance of genes encoding transmembrane receptors associated with immune cell activation and antigen presentation such as MHC class II molecules, Fc receptors, T cell co-stimulatory molecules (*CD80* and *CD83*), TLRs and Ig receptor superfamilies, and *TREMs* (**Figure 2D**). Although in mice CD163 is often referred to as a TAM marker, we did not observe a significant difference in CD163 expression between Br-RM and Br-TAM (**Figure S2B**). Comparison of DEGs between breast TEMo and Br-TAM showed minimal

overlap (**Figure 2E**). This indicates that the tissue environment determines the macrophage population's phenotype independent of their origin.

PCA and hierarchical clustering revealed distinct clusters of endometrial tissue resident macrophages (En-RM) and endometrial cancer TAMs (En-TAM) (**Figure 2F,G**). Limma DEA between En-RM and En-TAM identified 831 DEGs (115 up and 716 down; FDR <= 0.05, **File S1F**). Gene ontology analysis reported several enriched GO terms such as phagocytosis, immune response, cell communication, migration and blood vessel development (**Figure 2H**). Additionally, a number of genes encoding transmembrane receptors, soluble factors, transcription factors and enzymes were differentially expressed; the scavenger receptor *MARCO*, *TREM1*, *FCG2RB* and *IL21RG* were up-regulated in En-TAM as compared to En-RM (**Figure 2I**). Similar to that found in breast cancer, En-TAMs have minimal similarity to endometrial TEMo suggesting again that the cancer microenvironment directs the phenotype (**Figure 2J**).

To better understand TAMs in different cancer types, we compared the gene expression profiles of Br-TAM and En-TAM. PCA and hierarchical clustering revealed two distinct groups (**Figure S2C**) with very few DEGs commonly up- and down-regulated (18 genes up and 35 down, **Figure 2K, File S1G**), indicating that breast and endometrial cancers activate cancer tissue-specific transcriptional profiles in TAMs. Resident macrophages from endometrial and breast tissue also exhibited a distinct transcriptional profile confirming the diversity of tissue macrophage phenotypes in homeostatic states (**Figure S2D**).

Macrophages exhibit distinct phenotypes in response to environmental stimuli and have been classified into two alternative polarization states, referred to as 'M1' and 'M2' with the latter being immune suppressive and pro-tumoral (Martinez et al., 2006). To determine whether these polarization states exist within human En- and Br-TAM, we performed gene set enrichment analysis (GSEA) using the M1/M2 signature as proposed by Martinez et al. (**File S1H**). Neither Br- nor En- TAM showed a preferential enrichment for M2-associated genes supporting the idea that TAM phenotypes are much more complex and cannot be simply categorized into binary states (**Figure S2E, F**). Similarly, canonical markers for M2 that have been identified in mice, such as arginase, were minimally, and not differentially expressed in either Br- or En-TAM **(File S1E, F**).

TAM gene signature is enriched in aggressive breast cancer tumors

Increased density of TAMs has been associated with poor clinical outcomes in many human cancers (Yang et al., 2018). Importantly, studies using transcriptomic datasets have identified immune cell-specific gene sets to deconvolute the tumor microenvironment and its role in cancer progression

(Charoentong et al., 2017; Gentles et al., 2015). Taking advantage of a previously defined and validated compendium of immune cells (Bindea et al., 2013; Tamborero et al., 2018), we sought to identify a TAM-specific immune signature. We focused on Br-TAM as breast cancer has been the subject of greater number of in depth studies, and TAMs have been well characterized in mouse models of breast cancer. We selected up-regulated genes in Br-TAM compared to Br-RM ($Log_2FC > 3$, FDR <= 0.05) that were also highly co-expressed in the METABRIC cohort (Curtis et al., 2012), while filtering out genes belonging to other immune cell types (Tamborero et al., 2018) or those expressed by cancer cells (**File S1I**). As a result, we identified a 37-gene TAM signature (**File S1J**). We performed whole tumor RNA-seq on an independent cohort of 47 breast cancer patients (**DUKE cohort, see File S2 for clinical information**) and evaluated the expression of our TAM signature. The DUKE cohort samples (N = 37) were assigned to breast cancer molecular subtypes based on the PAM50 classification (Parker et al., 2009) with the TAM signature showing significantly higher expression in HER2 compared to Luminal A or B samples (p = 0.02) (**Figure S3A, B, File S3A**). Colony Stimulating Factor 1 (CSF1) is the major macrophage growth factor regulating their survival, differentiation and proliferation and its receptor is therefore a promising target for TAM depletion in cancer (Ries et al., 2014). A previous study of breast cancer, defined a 112-gene CSF1 response signature associated with higher tumor grade, decreased expression of estrogen and progesterone receptor and higher mutation rate (Beck et al., 2009). Using this CSF1-response signature, we stratified our dataset (DUKE cohort) into 3 groups (CSF1-High, -Mid and -Low). The TAM signature was significantly higher in the CSF1-High compared to -Mid and -Low groups suggesting that TAMs are associated with more aggressive tumors (**Figure 2L, File S3B**).

We next sought to investigate whether the identified TAM signature was associated with clinical outcome in the METABRIC cohort. As with the DUKE cohort, we observed a higher expression of the TAM signature in Basal, claudin-low, HER2 and Luminal B compared to Luminal A tumors, again suggesting an association of the TAM signature with more aggressive tumors (**Figure 2M, File S3C**). Consistent with these data, high expression of the TAM signature was significantly associated with shorter disease-specific survival (HR = 1.3, p = 0.006) (**Figure 2N**). A previously identified macrophage immune signature (Bindea et al., 2013; Tamborero et al., 2018), consisting mainly of lineage markers, showed a similar trend of high expression in aggressive tumors, but was not significantly associated with disease-specific survival (HR = 1.17, p = 0.1) (**Figure S3C, D, File S1I, File S3D)**. Taken together, these results suggest a positive association of unique populations of TAMs with poor clinical outcomes and more aggressive breast cancers.

Identification of breast TAM markers

One of the main limitations of targeting TAMs for therapeutic approaches is the lack of reliable and specific markers. To address this, we selected genes encoding transmembrane receptors in our TAM signature. We selected *SIGLEC1 (CD169)*, as it was one of the top up-regulated genes in Br-TAM (Log_2FC = 7.2, FDR = 0.0017) compared to Br-RM and it was also correlated with expression of the pan-macrophage marker CD163 (R = 0.62, p = 2.2e-16) (**Figure 3A**). In the METABRIC cohort, univariate analysis showed that SIGLEC1 high expression was significantly associated with shorter disease-specific survival (HR = 1.5, p = 1.2e-0.5, **Figure 3B**). Consistent with this, in Cox multivariate analysis after adjusting for clinical parameters such as ER, PR, HER2, grade and tumor size, SIGLEC1 high expression was independently significantly associated with shorter disease-specific survival (HR = 1.42, p = 1.85e-0.4, **File S3E**). Internal validation by qPCR confirmed the significant up-regulation of *SIGLEC1* mRNA observed in the RNA-seq analysis (**Figure 3C**). Furthermore, *SIGLEC1* showed significantly higher expression in breast tumor stroma compared to normal breast stroma in the two datasets available (**Figure 3D**) (Finak et al., 2008; Karnoub et al., 2007).

We used multicolour flow cytometric analysis to determine SIGLEC1 expression at the protein level in an independent cohort of breast cancer patients and found that SIGLEC1 was expressed on Br-TAM (**Figure S4A, B**), but not on other immune cells (**Figure S4C, D**) or CD45 negative non-immune cells (**Figure S4 E, F**), indicating specificity to macrophages/TAMs (**Figure S4, G, H**). In the circulation, classical and non-classical monocytes (**Figure S4 I, K**), but not granulocytes (**Figure S4 L, M**) exhibited low expression of SIGLEC1 with no difference between cancer and non-cancer patients observed. These data reveal preferential up-regulation of SIGLEC1 on Br-TAM.

Having established that SIGLEC1 is only significantly expressed by Br-TAM, we performed immunofluorescent (IF) staining using anti-SIGLEC1 and anti-CD163 antibodies on tissue biopsies from patients with invasive breast cancer and benign lesions (**Figure 3E**). Using machine-learning image analysis (**Figure S5A**) for unbiased quantification, we were able to segment and classify CD163 and SIGLEC1 single- and double-positive populations and determine their numbers within whole and sub-regions of the tissue sections. Cancer tissues had higher numbers of macrophages per mm^2 tissue area, and a higher percentage of SIGLEC1-positive cells compared to benign tissue (**Figure 3F**); results that were further confirmed by confocal microscopy of the stained sections (**Figure S5B**). These results indicate that SIGLEC1 is a human breast TAM-associated marker.

SIGLEC1 positive macrophages accumulate in Basal and HER2 breast cancers

To investigate expression of SIGLEC1 in different breast cancer subtypes we performed multiplex immunohistochemistry (Tsujikawa et al., 2017) on breast cancer tissues that had been independently acquired from the DUKE cohort. Using image cytometry we identified 3 distinct Br-TAM subtypes (CSFR1+CCR2-CD68+CD163+SIGLEC1-, CSFR1+CCR2-CD68+CD163+SIGLEC1+ and CSFR1+CCR2-CD68+CD163-SIGLEC1+, **Figure S6A**) confirming results reported in **Figure 3F**. Quantification of these three Br-TAM populations revealed enrichment in Basal tumors compared to HER2 and luminal subtypes, while the three subsets were almost absent in tissues from prophylactic mastectomies (**Figure S6B, C**). This is consistent with the increased expression of the TAM signature in aggressive breast tumors at the mRNA level.

Next, we investigated the regulation of SIGLEC1 expression in human macrophages using human monocyte-derived macrophages (MDM), induced pluripotent stem cell (iPSC) derived macrophages (iPSDM), and THP1 cells differentiated into macrophages using PMA (PMA-THP1). All three were exposed to conditioned medium (CM) from MDA-MB-231 and MDA-MB-468 cell line derived from triple negative breast cancers (Neve et al., 2006). CM from both cell lines increased expression of *SIGLEC1* mRNA in MDM and PMA-THP1 (**Figure S7A, B**). Additionally, CM enhanced SIGLEC1 protein expression on the cell surface of iPSDM (**Figure S7C, D**). These results indicated that cancer cells actively enhance the expression of SIGLEC1 on human macrophages.

In order to further investigate the stimulus generated by cancer cells we stimulated PMA-THP1 macrophages with a panel of pro-inflammatory and anti-inflammatory cytokines and measured SIGLEC1 expression by qPCR. The inflammatory mediator positive control, Lipopolysaccharides (LPS) and the pro-inflammatory cytokine Tumor Necrosis Factor α (TNFα) were the main modulators of SIGLEC1 expression, while Interleukin 1β (IL1β) and Interferon γ (IFNγ) produced a modest effect (**Figure S7E**). Conversely, anti-inflammatory cytokines did not affect SIGLEC1 expression in a significant way, except for a down-regulation after combined exposure with IL4 and TGFβ (**Figure S7F**).

In order to investigate if cancer cells produce TNFα, we performed ELISA on MDA-MB-231 and MDA-MB-468 CM, but did not detect significant levels of this cytokine (**Figure 3G**). In contrast, qPCR analysis indicated a significant up-regulation of *TNFα* at the mRNA level in Br-TAM compared to Br-RM (**Figure 3H**). Consistent with this elevated expression in Br-TAM, MDM and iPSDM incubated with either MDA-MB-231 or MDA-MB-468 CM produced significantly higher levels of TNFα compared to untreated controls at the protein level (**Figure 3G, I, File S4A**). We next neutralized TNFα in MDA-MB-231 and MDA-MB-468 CM-treated iPSDM (**Figure 3G**) and

exposed new iPSDM to the neutralized CM. TNFα neutralization resulted in a significant reduction of *SIGLEC1* expression compared to isotype control treated CM (**Figure 3J,K**). These results indicate that Br-TAM respond to cancer signals by up-regulating the expression of SIGLEC1 and by producing TNFα that further supports SIGLEC1 expression in macrophages.

CCL8 is a breast TAM marker

In order to identify additional mediators of the cross-talk between human cancer cells and macrophages we performed a qPCR array for inflammatory proteins on PMA-THP1 cells incubated with either MDA-MB-231 or MDA-MB-468 CM (**Figure 4A, B**). PMA-THP1 exposed to the different CMs commonly up-regulated the expression of 19 pro-inflammatory genes (**Figure 4C, File S4B**). By comparing this list with genes up-regulated in Br-TAM we identified seven factors commonly up-regulated (**Figure 4D**). Among these 7 factors *CCL8* (Monocyte Chemotactic protein-2 or MCP-2), was the most significantly up-regulated factor in the Br-TAM dataset. Interestingly, CCL8 has been reported to play a role in the tumor microenvironment by supporting mouse mammary cancer cells dissemination (Farmaki et al., 2016). In our data, CCL8 was correlated with CD163 expression (R = 0.68, p = 2.2e-16) (**Figure 4E**). In the METABRIC cohort, univariate analysis showed that CCL8 high expression was significantly associated with shorter disease-specific survival (HR = 1.3, p = 0.0019, **Figure 4F**). However, in Cox multivariate analysis, after adjusting for clinical parameters such as ER, PR, HER2, grade and tumor size, high CCL8 expression was not independently significantly associated with shorter disease-specific survival (HR = 1.16, p = 0.13, **File S3F**). Internal validation by qPCR on samples used for RNA-seq showed significant up-regulation of the *CCL8* transcript in Br-TAM (**Figure 4G**). We next validated by incubating PMA-THP1 macrophages, MDM and iPSDM with cancer CM and assessed CCL8 mRNA and protein levels (**Figure S7 G-I and Figure 4H**). In addition, *CCL8* fluorescence in situ hybridization (FISH) analysis of breast cancer tissue sections revealed that *CCL8* mRNA is found in Br-TAM but not in cancer cells (**Figure 4I**). There were no differences in CCL8 serum levels between normal and cancer patients indicating local production (**Figure S7J**). CCL8 production in human macrophages was induced by both pro-inflammatory and anti-inflammatory stimulation (**Figure S7K, L**) consistent with reports using cultured mouse macrophages (Makita et al., 2015).

Similarly to the observations with SIGLEC1, TNFα modulated the expression of CCL8 (**Figure S7K**). We neutralized TNFα in MDA-MB-231 and MDA-MB-468 CM-treated iPSDM with neutralizing antibodies and exposed new iPSDM to the neutralized CM. TNFα neutralization resulted in a significant reduction of *CCL8* expression compared to isotype control treated CM, confirming a role for TNFα in CCL8 regulation in macrophages exposed to cancer cell CM (**Figure 4J, K**). CCL8 treatment of both cancer cell lines significantly up-regulated the expression of CSF1, at both the

mRNA and protein level ($Log_2FC > 1$, $p < 0.05$, **Figure 4L, File S4C**), as well as TNFα and IL1β (**Figure 4M**). These results indicate that CCL8 is a marker of human breast TAMs and that cancer cells are able to modulate its expression in these cells by a TNFα-dependent mechanism. In turn, cancer cells respond to CCL8 by producing the macrophage survival and proliferation factor (CSF-1) and pro-inflammatory mediators, which further propagate the auto-stimulatory loop.

CCL8 enhances breast cancer cell motility and monocyte recruitment

We investigated the effect of CCL8 on cancer cells. Cancer cell lines were analyzed for expression of the five reported CCL8 receptors (**Figure S8A, B**). Of these CCR1, 2, 5 and 8 were detected on the cell surface of both MDA-MB-231 and MDA-MB-468 cells. We then stimulated MDA-MB-231 and MDA-MB-468 with recombinant CCL8 (rCCL8) and performed a qPCR array for genes associated with breast cancer progression. Using stringent criteria for changes in gene expression ($Log_2FC >2$, $p < 0.05$), six genes were identified that were commonly up-regulated in both cell lines following stimulation with rCCL8 (**Figure S9A-C**). These gene products have been predicted to be involved in cancer cell invasion (*MMP2, MMP9, ADAM23*) (Roomi et al., 2009) and progression (*IL-6, EGF, GLI1*) (Knüpfer and Preiß, 2006; Masuda et al., 2012) (**Figure S9D, File S4C**). Similar genes were identified by a metastasis qPCR array after exposure of MDA-MB-231 and MDA-MB-468 with CM from cancer-cell primed MDM (**Figure S9E-H, File S4D**). Consistent with the up-regulated expression of genes involved in invasion, rCCL8 treatment enhanced motility of MDA-MB-231 cells (**Figure 4N, O**), and to a greater extent than previously reported for CCL2 (Bin Fang et al., 2012). Stimulation with rCCL8 did not affect cell proliferation of either breast cancer cell line (**Figure S8C**). Finally, we assessed the ability of CCL8 to recruit monocytes using an *in vitro* chemotaxis assay with THP1 monocytic cells in the presence of CCL2 and CCL8 as chemo-attractants. Both CCL2 and CCL8 attract these monocytic cells compared to control, indicating a similar effect of these chemokines on myeloid cell recruitment (**Figure 4P**). These data indicate a potential CCL8-dependent cross-talk between human macrophages and neoplastic cells that drives tumor cell invasion and monocyte infiltration.

SIGLEC1/CCL8 gene signature is an independent prognostic factor in ER+ breast cancer

To assess whether a 2-gene *SIGLEC1/CCL8* signature had clinical relevance in breast cancer, Cox proportional hazard regression analysis was performed on a breast cancer stroma dataset (Finak et al., 2008) representing 53 patients suffering 17 recurrence events reported over a median follow-up time of 8.7 years. Gene expression values of *SIGLEC1/CCL8* were dichotomized into high and low expression groups according to all possible cutoffs (Pearce et al., 2017). Univariate analysis revealed

that *SIGLEC1/CCL8* high expression was associated with shorter recurrence-free survival (RFS) (HR = 3.3, p = 0.018) (**Figure 5A**). To further validate the clinical relevance of the *SIGLEC1/CCL8* gene signature, we utilized the METABRIC cohort (N = 1350) with 456 breast-cancer specific events over a median follow-up time of 9.69 years. In univariate analysis high expression of *SIGLEC1/CCL8* was significantly associated with shorter disease-specific survival (HR = 1.4, p = 4e-04) (**Figure 5B**), along with HER2 status (HR = 2.1, p = 2.3e-10), grade (HR = 1.8, p = 1.9e-09) and tumor size (HR =1.8, p = 0.002). Conversely, ER (HR = 0.6, p = 7.8e-07) and PR (HR = 0.64, p = 3e-06) status were significantly associated with better disease-specific survival. In Cox multivariate analysis, *SIGLEC1/CCL8* high expression was associated with shorter disease-specific survival but didn't reach significance (HR = 1.2, p = 0.06, **File S5**).

In a subset of ER positive patients from the METABRIC cohort (N = 960), univariate analysis revealed that *SIGLEC1/CCL8* high expression was significantly associated with shorter disease-specific survival (HR = 1.5, p = 0.001), along with grade (HR = 1.7, p = 9e-06) and age (HR = 1.5, p = 0.002) (**Figure 5C, File S5**). Cox multivariate analysis demonstrated that *SIGLEC1/CCL8* high expression was independently significantly associated with shorter disease-specific survival (HR = 1.35, p = 0.014) along with, grade (HR = 1.54, p = 3.4e-04) and age (HR = 1.44, p = 0.008).

Discussion

In mouse models of cancer, cells of the mononuclear phagocytic system play profound roles in shaping the tumour microenvironment to one that promotes malignancy. Despite the large number of studies in mice and clinical correlative data suggestive of similar roles in human cancers for TAMs, there is little data describing their phenotypes. In the present work, we have shown that in breast and endometrial cancers circulating monocytes and TAMs respond to the presence of malignancy by altering their transcriptomes and therefore showing distinct profiles compared to healthy women. TAM transcriptional profiling yielded a 37-gene TAM-specific signature that is highly expressed in aggressive forms of breast cancers and associates with poor clinical outcomes. Furthermore, detailed analysis and mechanistic studies showed paracrine signalling interactions between tumor cells and TAMs that involve SIGLEC1 and CCL8, two TAM markers that in multivariate analysis were independent predictors of disease-specific survival in ER positive patients.

We observed an expansion in the non-classical monocytic population $CD14^{+}CD16^{++}$, a result consistent with evidence for a role of these cells in cancer (Bharat et al., 2017). In contrast, Hanna et al. reported that in mouse models of metastasis depletion of non-classical monocytes correlated with enhanced metastasis through inhibition of NK cell activity (Hanna et al., 2015). Recently however, non-classical monocytes have been shown to contribute to anti-VEGF therapy resistance in mouse models of cancer (Jung et al., 2017). This therapy has also been associated with enhanced CX3CL1

levels in human colon cancers, leading to the recruitment of non-classical monocytes to the vascular bed of the tumor, where they promote accumulation of neutrophils and immune suppression through IL-10 secretion (Jung et al., 2017). In our patient cohort, we detected significantly higher levels of CX3CL1 but not CCL2 in the sera of breast cancer patients compared to healthy controls. This chemokine increase could explain the elevated number of monocytes, and also potential activation when this cancer is present. In order to rule out the possibility that the transcriptional differences observed in total monocyte populations were only due to alteration of subpopulation frequency, we sequenced equal amount of non-classical monocytes from normal and cancer patients. The two populations were still very different suggesting that cancer was the main driver of the TEMo transcriptional shift. Using total monocyte transcriptional profiles we identified a 17-gene signature, which indicated the presence of cancer. Such an approach could be translatable to the clinic where simple monocyte purification methods coupled with array technology could be employed. However, before this method would be applicable a large multicenter trial is needed.

In mouse models of cancer, monocytes are recruited to primary or metastatic tumors where they differentiate to TAMs that promote tumor progression and metastasis (Arwert et al., 2018). However, little is known about TAMs in human cancers. We profiled TAMs from samples belonging to the same cohort of patients from which we isolated TEMos and defined their transcriptional landscape through bulk RNA-seq. Surprisingly, in contrast to the monocytes, TAM transcriptomes from endometrial and breast cancers are distinct from each other, from their respective resident macrophages and their progenitor monocytes. These data suggest the existence of cancer specific niches at the tissue level that influence the TAM transcriptional profile according to tumor location and subtype. Multiplex IHC identified at least 3 TAM sub-populations (CSFR1+CCR2-CD68+CD163+SIGLEC1-, CSFR1+CCR2-CD68+CD163+SIGLEC1+ and CSFR1+CCR2-CD68+CD163-SIGLEC1+) in breast cancer patient samples and showed that there is considerable heterogeneity in TAM populations within the tumor. Although macrophages were traditionally classified into M1 and M2 polarization states, more recent studies describe a spectrum of activation states (Xue et al., 2014) and an association of both states with tumor progression (Franklin et al., 2014). Indeed, our analysis failed to reveal a unique polarization state in human TAMs. A recent study that analysed the transcriptomes of CD45+ cells from healthy tissue and breast tumors at the single cell level (Azizi et al., 2018), confirmed the heterogeneity of myeloid cells. However, unlike this study the moderate depth of scRNA-seq studies doesn't allow for identification of novel markers.

High expression of macrophage gene signatures has been associated with high tumor grade and poor clinical outcomes (Gentles et al., 2015). In our study, we identified a TAM signature that was highly expressed in the most aggressive breast cancer subtypes and enriched in a CSF-1-high group that has been previously associated with higher tumor grade, decreased expression of estrogen and

progesterone receptor, and higher mutation rate (Beck et al., 2009). Compared to a pan-macrophage immune signature (Bindea et al., 2013; Tamborero et al., 2018), the TAM signature was associated with shorter disease-specific survival in the METABRIC cohort. These results, along with recent evidence of the role of TAMs in chemo- and immuno-therapy resistance (Neubert et al., 2018), highlight the need to study TAMs in human cancers and to identify markers for TAM-specific targeting. Therefore, we focused on transmembrane receptors included in the TAM signature, of which, SIGLEC1, a sialic binding receptor mainly expressed by macrophages, was the most highly differentially expressed in our Br-TAM dataset as well as in publically available breast cancer stroma datasets. In homeostatic conditions, SIGLEC1 positive macrophages are mainly localized in the bone marrow, liver, spleen, colon and lymph node (Chávez-Galán et al., 2015) and they are thought to be involved in erythropoiesis and modulation of adaptive immune responses (Chávez-Galán et al., 2015). Consistent with our findings, SIGLEC1 positive macrophages have been identified in colorectal (Li et al., 2015) and hepatocellular carcinoma (Zhang et al., 2016). Interestingly, while infiltration of SIGLEC1 positive macrophages in colorectal cancer was associated with tumor progression, in hepatocellular carcinoma they predicted favourable patient outcomes, underpinning the hypothesis that TAM phenotypes/activation are organ- and cancer-specific (Zhang et al., 2016).

In order to elucidate the crosstalk between human TAMs and breast cancer cells, we focussed on soluble factors produced by TAMs in response to cancer cell CM. Our screening identified CCL8 as the top upregulated soluble factor in Br-TAM. This chemokine is involved in the regulation of activation of immune cells involved in inflammatory responses (Proost et al., 1996). CCL8 expression was reported to be associated with metastasis formation in melanoma (Barbai et al., 2015) and more relevantly in a mouse model of breast cancer (Farmaki et al., 2016), where CCL8 promoted cancer cell invasion and motility. Interestingly, SIGLEC1 positive macrophages in the mouse intestine produce high levels of CCL8 in response to sterile and non-sterile inflammatory stimuli (Asano et al., 2015). CCL8 production was also shown to sustain dextran sulphate sodium colitis and to recruit pro-inflammatory monocytes to the inflamed site. We demonstrated that TAMs are the major source of CCL8, and CCL8 and SIGLEC1 engage in a tumor cell-TAM regulatory loop, involving TNFα that in turn enhances their expression and leads to increased tumor cell motility. Our data showed that cancer cells and TAMs secrete high levels of TNFα that further supports CCL8 production in the tumor microenvironment and that cancer cells respond to the presence of CCL8 by producing significant higher levels of CSF-1, the major survival and proliferation factor for macrophages. The high concentration of CCL8 not only supports the cancer-TAM crosstalk, but also acts as a monocyte chemoattractant. We therefore propose CCL8 will also increase monocyte infiltration into the tumor site, thus generating more pro-tumoral TAMs (**Figure 6**). Consistent with this positive auto-regulatory loop that potentially increases TAM recruitment and cancer cell malignancy SIGLEC1 and CCL8 were associated with shorter disease-specific survival and recurrence-free survival in public datasets

derived from whole tumor homogenates. Such data reinforce the concept derived from mouse models that TAMs in the tumor microenvironment promote malignancy, and the identification of uniquely expressed genes in human TAMs provides opportunities for new therapeutic targets and diagnostic/prognostic markers.

Acknowledgements

This research was supported by Wellcome Trust (101067/Z/13/Z), MRC Centre grant MR/N022556/1, NIH grant PO1 CA100324 to JWP, Defense Breast Cancer Research Program (W81XWH-111-0702) to LMC/JWP, Susan G Komen Breast Cancer Foundation (KG111084 and KG110560) to ESH and LMC, from Breast Cancer Now to AHS and JMD, from CONACYT to MLY and Wellcome Trust (102610) to AS. The work was supported extensively by the Edinburgh Breast Unit team and particularly by Lorna Renshaw and Jane Keys in this unit and the Departments of Gynecological and Surgical Oncology at the Montefiore Medical Center. We would like to thank the CIR blood resource (AMREC #15-HV-013) for the rectuiment of blood from normal controls. We thank all the patients and volunteers who contributed to this study as well as all the clinical support teams.

Authors Contributions

LC, SF, LMC and JWP contributed to experiment conception and design, acquisition, analysis and interpretation of data as well as article drafting. DS, EL and TC performed all imaging acquisitions and image analysis. LF, HZ, AS, TC, AF, MM, MLY, MRM and ZA performed parts of the experiments reported in the paper. LMF supplied iPSC cells and methods. LW, EHS, HAS, AU, JMD provided clinical samples of endometrium and breast cancers and supporting clinical data and analysis. SF performed bioinformatics analysis of the data. AHS supervised the bioinformatics analysis and contributed novel methods. PA and PTS performed bioinformatics analysis on the DUKE cohort. All authors approved the article.

Conflict of interest

Patent protection is being sought for the prognostic and diagnostic signatures reported in this paper.

Material and Methods

Patient and control samples

All study protocols were approved by the IRB of the Albert Einstein Medical College (Bronx, NY, USA) and by The University of Edinburgh (Edinburgh, UK), and Duke University (Durham, NC) ethics committees as appropriate. Informed consent was obtained from all human subjects included in this study. For control samples, mononuclear cells were isolated from peripheral blood obtained from female healthy individuals through the New York Blood Center, USA or Cambridge Biosciences, UK. In some cases, blood was also donated from volunteers in the Bronx, NY who were age and weight matched to the Bronx cancer cohort. Peripheral blood (20ml) was obtained from breast and endometrial cancer patients attending the Montefiore Medical Center, Bronx, NY, USA and breast cancer patients (15mL) from NHS, Edinburgh, Scotland, UK. Breast cancer tissue (0.1-1 grams) and endometrial cancer tissue (0.1-1 grams) was obtained from Montefiore Medical Center, NY, USA, the NHS, Edinburgh, Scotland, UK (breast only), and Duke University, Durham NC, USA (breast only). Pathologically the breast cancer patients consisted of invasive breast cancers. The endometrial cancer patients consisted of Type I (Endocarcinoma) and Type II (UPSC) cancers. The Duke cohort patients had biopsy-confirmed invasive tumors at least 1.5 cm at diagnosis (**see File S2 for detailed clinical information**) Normal breast tissue from mammoplasty reduction surgeries (25-50 grams) was obtained from the Human Tissue Procurement Facility (HTPF), Ohio State University, USA; normal/benign endometrial tissue (1-2 grams) was obtained after surgery for conditions unrelated to cancer from Montefiore Medical Center, NY, USA; breast tissue (0.5-1 grams) from patients with benign conditions was obtained from NHS, Edinburgh, Scotland, UK. The exclusion criteria for cancer patients at baseline included systemic metastatic disease, any inflammatory disorder, and active infection or immunocompromised status not related to cancer. All the patients recruited were chemotherapy and radiotherapy naive before collection. Blood collection was performed before surgery. Tissue samples were collected as part of definitive surgical intervention. For the DUKE cohort, patients with newly diagnosed invasive breast cancer were recruited from the breast clinic at the Duke Cancer Institute from 2012 to 2016. Eligible patients had no evidence of metastatic disease and were scheduled to undergo either lumpectomy or mastectomy. Primary tumors measured at least 1.5cm clinically. Patients with either node negative or node positive disease were included.Blood was collected preoperatively, and trained pathologists collected tumor from surgical excision specimens. Both blood and tumor samples were shipped on ice to Oregon Health & Science University Hospital (OHSU) for immune and genomic assays.

Isolation of human blood monocytes

All control and cancer blood samples were collected and processed by the same person according to site, HZ in the Bronx and LC Edinburgh. They were processed as attained and not batched together according to sample type. All the blood samples were collected in Venous Blood Collection Tubes containing EDTA and stored immediately at 4°C after collection. Blood was centrifuged at 700 RCF for 10 min at 4°C in a swinging bucket rotor to separate cells from plasma, and was then subjected to centrifugation in conical tubes for 10 min at 16,000 x g at 4°C in a fixed angled rotor, immediately aliquoted and stored at -80°C. After red blood cell lysis (RBC lysis buffer, Biolegend) cells were centrifuged 500 RCF 5 min at 4°C, counted and stained for FACS analysis; the remaining cells were fresh or frozen in 10% v/v DMSO, 90% v/v Fetal bovine serum solution for subsequent cell sorting and RNA extraction.

Isolation of human tissue macrophages

Cancer tissue and normal endometrial tissue were washed with PBS in a petri dish and tissue was chopped into small fragments with a razorblade. The sample was transferred to a 15-50 ml tube according to size and Liberase enzymes TL and DL (Roche) and DNAse (Roche) were added. Tissue was digested at 37°C on a rotating wheel for 1h-18h depending on tissue weight; at the end of digestion the cell suspension was filtered and washed with PBS 1% w/v Bovine Serum Albumin (BSA, Sigma-Aldrich). The pellet was resuspended in PBS, 1%w/v BSA and cells counted and stained for FACS sorting or analysis. Macrophages were isolated according to the method described (Cassetta et al., 2016).

FACS Sorting and analysis

PBMCs or total blood cells were counted and resuspended in PBS, 1%w/v BSA; blocking of Fc receptors was performed by incubating samples with 10% v/v human serum (Sigma Aldrich) for 1h on ice. For FACS analysis $5x10^5$ cells were stained in a final volume of 100 µL using the following antibodies: CD45-PE-Texas Red, CD3-, CD56-, CD19-BV711, CD11b-BV605, CD14-BV510, CD16-EF450, CX3CR1-FITC, HLA-DR-BV650, CD64-APC-CY7, CD80-PECY7, CCR2-PECY7, CD86-APC, CD95-PECY7 (Biolegend). For monocyte and macrophage sorting cells were stained and antibody concentration was scaled up based on cell number; cells were stained with the following antibodies: CD45-AlexaFluor 700, CD3-, CD56-, CD19- PECY5, CD14-FITC, CD11b- PECy7, CD16- PE-Texas Red, CD163-APC (Biolegend). Cancer cell lines were stained for the 5 CCL8 receptors with the following antibodies: CCR1-PE, CCR2-PECy7, CCR3-FITC, CCR5-PE, CCR8-PE (Biolegend) (Full list **File S7)**. Cells were incubated in the dark for 1h on ice; after washing with PBS 1%w/v BSA (analysis) or PBS 0.1%w/v BSA (sorting) cells were filtered and resuspended in the

appropriate buffer before analysis or sorting. FACS analysis was performed using a 6-laser Fortessa flow cytometer (BD); FACS sorting was performed using FACS AriaII and FACS Fusion sorters (BD). Cell sorting was performed at 4ºC in 1.5 ml RNAse and DNAse free tubes (Simport, Canada) pre-filled with 750 µl of PBS 0.1%w/v BSA; at the end of each isolation a sorting purity check was performed. A minimum of 5,000 events in the monocyte/macrophage gate was acquired for FACS analysis. Results were analyzed with Flowjo (Treestar) or DIVA software (BD).

RNA extraction and sequencing of purified cells

Immediately after sorting all the samples were centrifuged at 450 RCF for 10 min at 4°C. The cell pellet was resuspended in 350 µL of RLT lysis buffer + 1% βmercaptoethanol, and RNA extracted with RNAeasy Microkit (Qiagen) according to manufacturer's instructions. RNA quantity was determined by QUBIT (Invitrogen); total RNA integrity was assessed by Agilent Bioanalyzer and the RNA Integrity Number (RIN) was calculated; samples that had a RIN > 7 were selected for RNA amplification and sequencing. RNA was amplified with Ovation RNA-seq Amplification kit v2 (Nugen) according to manufacturer's instructions. Amplified RNA was sent to Albert Einstein Genomic Facility or BGI where library preparation, fragmentation and paired-end multiplex sequencing were performed (HIseq 2000 and 2005, Illumina). All samples were processed and randomly assigned to lanes without knowledge of clinical identity to avoid bias and batch effects.

Sequencing alignment and Quantification

FastQ files of 2x100bp paired-end reads were quality checked using FastQC (Andrews, 2012). Samples were filtered for low quality reads (Phred score >= 20) and adapters were removed when necessary using Cutadapt. Quality controlled reads were then aligned to the Human reference genome (GRCh37/hg19) using STAR aligner (version 2.3) (Dobin et al., 2012). Quantification of genes was performed using the count function of HTSeq (Anders et al., 2014). Reads were counted at the gene level and the unstranded option was used (-s no). A summary of number of reads and % of alignment of all samples can be found in **File S1A**.

Statistical analysis for differentially expressed genes

All statistical calculations were performed in R programming language (version 3.2.3). For macrophage samples, genes with count per million (CPM) reads > 1 in at least N samples (N number of the fewest replicates of a condition) was retained. Gene expression levels were normalized using the Trimmed Mean of M-values (TMM) method using the calcNormFactors() function and $\log_2$ transformed using the cpm() function from the EdgeR package in R (Robinson and Oshlack, 2010).

Differential expression analysis was performed with sample quality weights using the package limma-voom package in R (Ritchie et al., 2015). For monocyte samples, gene expression levels were normalized according to upper-quantile normalization using calcNormFactors() function and $\log_2$ transformed using the cpm() function. Normalized reads were fed to the Combat function of the Surrogate Variable Analysis (SVA) package (Leek, 2014). Samples were corrected on the basis of the sequencing batch they were processed. Differential expression analysis was performed using the limma package. Significantly differentially expressed genes (DEGs) were selected with controlled False Positive Rate (B&H method) at 5% (FDR <= 0.05). Up-regulated genes were selected at a minimum $\log_2$ fold change of 1.5 and down-regulated genes at a minimum $\log_2$ fold change of -1.5. PCA plots were drawn using the TMM/$\log_2$ transformed (macrophages) or the batch effect corrected (monocytes) normalised values on expressed genes. Heatmaps were drawn on the normalized expression matrix using the pheatmap package in R. Euclidean distance and complete linkage were used for hierarchical clustering. Venn diagrams were constructed based on the overlapping differentially expressed transcripts (FDR <= 0.05, Log_2FC more or less than 1.5/-1.5).

Enrichment and Pathway Analysis

Gene set enrichment analysis was performed using the gsea() function from phenoTest package in R (Planet, 2013). The function is used to compute the enrichment scores and simulated enrichment scores for each variable and signature. For our analysis, the logscale variable was set to false, as the $\log_2$ transformed expression values were fed into the function and 20,000 simulations were used (B = 20,000). The Database for Annotation, Visualization and Integrated Discovery (DAVID) functional annotation tool was used for gene ontology and pathway (KEGG and Reactome) analysis on the list of differentially expressed genes (FDR <= 0.05, Log_2FC more or less than 1.5/-1.5). Important GO terms and pathways were selected based on an FDR <= 0.05.

Recursive feature elimination with Random forest (RFE-RF)

Recursive feature elimination (RFE) is a wrapper feature selection method that starts by fitting the classification model to all features and then each feature is ranked for its importance to the model. We assume that S is a sequence of values that indicates the number of features to be retained. At each iteration, the S_i top ranked features are used to refit the model and the subset of genes with the highest accuracy is selected. We used the RFE feature selection with resampling and a Random forest (RF) classifier as implemented in the caret package in Bioconductor (Kuhn, 2015). In brief, RF is an ensemble learning method that constructs many decision trees (forest) and uses a majority vote to make predictions (Breiman, 2001). Assuming N number of samples in the training set, the algorithm creates a subset of the data of the N samples with replacement, and at each node, *m* number of genes

are selected at random from all the genes M. The variable *m* that gives the best split is selected and used for a binary split. This procedure is repeated for each node until the tree is grown to terminal nodes k.

Before model training, the breast TEMo dataset of N = 77 samples (32 healthy, 45 breast cancer) was filtered for lowly expressed genes (CPM > 1 across conditions), normalized using upper-quantile normalization, $\log_2$ transformed using the *cpm()* function and corrected for batch effects using ComBat. Then, the dataset was split into training (70%, N = 55, 32 healthy, 23 cancer samples) and testing (30%, N = 22, 13 healthy, 9 cancer samples). While the test set was kept on the side, the training set was used for RFE-RF model training and feature selection using 5 times repeated 10-fold cross-validation (CV). In short, the training samples were randomly partitioned into k (k = 10) subsamples. Out of the k sub-samples, one is kept for testing the classifier, and the remaining k-1 subsamples are used for training the classifier. Then the whole process is repeated n= 5 times. For the training set, overall accuracy, sensitivity and specificity were calculated averaging the CV predictions for the optimal subset. The model with the highest average accuracy was selected as optimal. The optimal model was the fitted on the test set that had not been used for training or feature selection. For the RFE-RF model, we selected subsets of features ranging between 2 to 30, variable *ntree* was set to 500 trees and variable *mtry* was calculated as $\sqrt{p}$, where p is the number of features used during training the model. The receiver-operating characteristic (ROC) curves were drawn using the ROCR package in R (Sing et al., 2005). To determine the accuracy rates of the classifiers that can be obtained by chance, a RF model using the 17 selected genes was trained with permuted class labels. This process was performed during training 1,000 times using 5 x 10-fold CV. The accuracy of the 1,000 random classifiers was recorded. The P-value was calculated by counting the accuracy of the random classifiers that achieved similar or higher total accuracy compared to the observed accuracy of the RFE-RF model on the training data.

Publically available datasets

The following publically available datasets were used in this study:

a) Karnoub et al. (GSE8977) (Karnoub et al., 2007): total of 22 samples coming from breast ductal carcinoma-in-situ (DCIS) patients (N = 15) and invasive ductal (IDC) breast cancer patients (N = 7) were downloaded from GEO. Samples we processed and normalised using the robust Multi-Array average expression measure (RMA) from the affy package in R. Probes representing the same gene were averaged to a single value.
b) Finak et al. (GSE9014) (Finak et al., 2008): total of 59 samples coming from breast cancer stroma patients (N = 53) and healthy controls (N = 6) including updated clinical information were downloaded from GEO. Technical replicates were averaged to a single array using the

averarrays() function from limma package in R. Data were then quantile normalised using the normalizeQuantiles() function. Samples were annotated and probes representing the same gene were averaged to a single value.

c) METABRIC cohort (Curtis et al., 2012): Microarray gene expression data and associated clinical information (N = 1980) (Log_2 transformed intensity values) were downloaded from the cBioPortal for cancer genomics database (http://www.cbioportal.org/) under the study name Breast cancer. Gene expression values were quantile normalized and samples with gene expression and corresponding clinical information were selected resulting in N = 1353 patients. Data were filtered for missing values and samples with molecular subtype NC were removed. The filtering resulted in N = 1350 patients that were used for further analysis. For survival analysis, events were censored based on disease-related deaths (Died of disease = 1; Living or Died of other causes = 0).

d) Cancer cell Encyclopedia (CCLE) data: Gene expression RPKM normalized reads from breast cancer cell lines (N = 57) were downloaded from the CCLE website (https://portals.broadinstitute.org/ccle).

TAM signature

As a starting point for the selection of the immune signature we used the up-regulated genes in Br-TAM compared to Br-RM (N = 553, $Log_2FC > 3$ and FDR < 0.05). This gene list of highly differentially expressed genes was filtered using the compendium of immune genes that includes 17 immune cell-specific gene sets as initially assembled by Bindea et al. and Charoentong et al. (Bindea et al., 2013; Charoentong et al., 2017) and most recently validated by Tamborero et al. (Tamborero et al., 2018) (**File S1I**). After filtering, N = 528 TAM related genes were selected. In order to identify the most relevant genes we used the METABRIC cohort and correlation analysis. Genes were considered coexpressed when having an absolute Pearson correlation of R >= 0.5 (findCorrelation() function from the caret package in R (Kuhn, 2015)). This threshold was selected in order to satisfy two main aims: a) genes with relatively high correlation would not be considered a chance event; b) selection of a relatively small number of genes in order to be suitable for gene set enrichment and survival analysis (Charoentong et al., 2017). Additionally, genes were selected based on their positive Pearson correlation (R > 0.5, p <= 0.05) with known TAM marker CD163, resulting in N = 37 genes (**File S1J**). Finally, we downloaded breast cancer cell line data from the CCLE database (N = 57 breast cancer cell lines) in order to filter out genes expressed by tumor cells. Genes with median expression of $Log_2RPKM > 6$ were considered expressed in tumor cells (TAM: median = 0.031). This resulted in a set of 37 genes expressed by TAMs and not tumor cells or other immune-specific signatures (**File S1J**). The TAM and macrophage signature scores were calculated as the median of

expression of the TAM or macrophage signature genes (Tamborero et al., 2018) using the median centered normalised values.

Survival analysis

For the TAM signature and SIGLEC1/CCL8 signature the summed normalised gene expression values were dichotomized based on the optimal cutoff calculated by iteratively calculating every possible expression cutoff (n-1) and selecting the value with the lowest p-value (Pearce et al., 2017). For the METABRIC cohort, disease-specific survival (DSS) was used as an endpoint. For the breast cancer stroma dataset, recurrence-free survival (RFS) was used as an endpoint and censored at date of last follow-up. Survival curves were estimated using the Kaplan Meier method (survival and survminer R packages). For SIGLEC1 and CCL8 single gene survival analysis, clinical risk factors such as ER status (positive or negative), PR status (positive or negative), HER2 status (positive or negative), histological grade (I,II or III), age (greater or less than 55) and tumor size (greater or less than 50mm) were used in the univariate and multivariate models. Candidate prognostic factors for RFS and DSS with a p-value (Wald test) lower than 0.05 in univariate analysis were used in the multivariate analysis. Multivariate analysis was performed by fitting a Cox proportional hazard regression model. The Cox regression model was used to calculate the Hazard ratio (HR) and 95% confidence internal (CI). A p-value less than 0.05 based on a Wald test was considered significant.

RNA-seq of total tissue breast cancer

RNA isolated from 47 breast cancer tumors (DUKE cohort) was utilized for RNA-seq. These RNA samples were converted into a library of cDNA fragments. Illumina sequencing adapters were added and 50 bp single end read sequence was obtained using Illumina HiSeq. Quality check was performed on these sequence reads using FastQC. PCR primers and adapters were filtered out of the sequence reads using Trimmomatic (Bolger et al., 2014). Filtered reads were aligned to reference genome build hg19 using TopHat 2.0.12, a splice junction aligner (Kim et al., 2013). Aligned sequences were assembled into transcripts. Transcript abundance was estimated as Fragments Per Kilobase of exon per Million fragments mapped (FPKM), using Cufflinks 2.2.1 (Roberts et al., 2011; Trapnell et al., 2010). FPKM estimates were normalized using Cuffnorm. Further data was quartile normalized and batch effects were removed using ComBat (Johnson et al., 2007). Samples were classified into CSF1 High, Mid and Low expression groups using K-means clustering on the expression of CSF1 signature genes (Beck et al., 2009). TAM signature score was estimated from the median of expression of TAM signature genes. Samples were assigned to breast cancer subtypes based on hierarchical clustering of PAM50 genes (Parker et al., 2009). Clustering was performed using R package pheatmap_1.0.8. Correlation was used as a distance measure and average was used as clustering method.

Quantitative PCR

Cells were lysed and RNA extracted with RNAeasy Microkit (Qiagen) according to manufacturer's instructions. Typically, 0.1 ug of total RNA was reverse transcribed using Super Script Vilo kit (Invitrogen) and the cDNA generated was used for semi quantitative PCR on a 7900 Real Time cycler (Applied Biosystem) as per manufacturer's instructions. Target gene expression was normalized to the expression of the housekeeping gene GAPDH. Relative gene expression was calculated using the standard $2^{-\Delta\Delta CT}$ method. Primers were designed using Primer Bank (Wang et al., 2012). The full list of primers used can be found in **File S6A.**

Immunofluorescence and quantitation

All tissues were fixed in 4% w/v paraformaldehyde, dehydrated and embedded in paraffin blocks; 5 µm sections were cut onto positively charged glass slides and stained with the following antibodies: CD163 (Novocastra NCL-LCD163, Clone 10D6) dilution 1:1000, SIGLEC1 (Novus Biologicals, NPB2-309003, polyclonal) dilution 1:100. High throughput immunofluorescence was performed by the SURF Facility at the University of Edinburgh (http://surf.ed.ac.uk/facilities/immunodetection-and-histological-imaging/) after primary antibody optimization. Immunofluorescently stained tissues were batch-scanned on a Zeiss AxioScan.Z1 (Carl Zeiss, Oberkochen, Germany) with specific scan profiles for each stain group and using a 40x Plan-Apochromat 0.95NA coverslip corrected air objective. Slide scanned images were imported into a Definiens Tissue Studio workspace (Definiens AG, Munich, Germany) and pre-processed for nuclear detection and cell simulation using built-in nuclear detection and cell growth algorithms. The pre-processed workspace was then imported into Definiens Developer XD (Definiens AG, Munich, Germany) for further processing, quality control, machine learning, and k-Nearest Neighbour classification and output compiled in Mathematica 10.3 (Wolfram Inc., Champaign, Illinois, United States) and tabulated in a spreadsheet (**File S6B**). Incomplete or low-quality nuclei and cells were discarded using a combination of DAPI pixel intensities and standard deviation. For CD163, examples of 300 cells each were given for positive and negative cases in a single large tissue sub-region of one cancer tissue previously identified to show the most variation of intensity. These class samples were used to optimize a feature space consisting of 49 subjectively selected morphological, textural, statistical, and intensity-based metrics. Feature space optimization indicated 19 features as being most important for separation of both populations using a Euclidean distance matrix. A classifier algorithm was used to compile these 19 metrics for each given class sample in each population and then used to classify all remaining cells in that tissue. A selection of at least 10 incorrectly classed cells were then manually corrected and added to the relevant class sample populations before recompiling the 19-dimensional feature space and reclassifying the whole tissue.

This iterative learning process was repeated at least 10 times with a final sample size of 400-500 cells for each class. The classifier was stored as .xml and used to batch classify the entire data set of tissues from all patients.

Multiplex immunohistochemistry

Staining: Formalin-fixed paraffin embedded (FFPE) tissue samples (5 μm) were used for iterative multiplex immunohistochemistry as previously described (Tsujikawa et al., 2017). Briefly, following standard antigen retrieval and blocking, primary antibodies (listed in table below) were applied to tissue sections and incubated overnight at 4°C. Primary antibodies were detected using a species-specific F(ab') fragment-specific secondary antibody-labeled polymer-based peroxidase system (Histofine, Nichirei Biosciences Inc, Japan) in conjunction with 3-amino-9-ethylcarbazole (AEC). Images were acquired using Aperio ImageScope AT (Leica Biosystems) and slides were subject to iterative cycles of staining.

Image processing and analysis: All image processing and analysis was performed as previously described (Tsujikawa et al., 2017) on three regions/slide, which encompassed the total tissue area. Image cytometry was performed using FCS Express 5 Image Cytometry (De Novo Software) and cell populations were determined using multiparameter cytometric image analysis (see gating schema). Cell populations were normalized to total cell number (Cells/Total Cells) and populations were quantified. Unsupervised hierarchical clustering was performed using R package pheatmap_1.0.8. Correlation was used as a distance measure and average was used as clustering method (Tsujikawa et al., 2017).

Antibodies used for multiplex IHC

Target	Company/Product#	Clone	Dilution
CD45	eBiosciences: 14-0459-82	H130	1:100
CD3	ThermoFisher Scientific:MA1-90582	SP7	1:150
CD8	ThermoFisher Scientific:MA5-13473	C8/144B	1:100
CSF1R	abcam: ab183316	SP211	1:150
CD169	Millipore: MABT328	5F1.1	1:200
CD163	ThermoFisher: MA5-11458	10D6	1:100
CD56	Santa Cruz Biotech	123C3	1:100
CCR2	Novus: MAB150-SP	48607	1:400

Monocyte Derived Macrophages (MDM) isolation and stimulation

Peripheral blood mononuclear cells (PBMC) were extracted after Ficoll gradient stratification and centrifugation as previously described (Cassetta et al., 2013). PBMC were seeded in a 12-well plate (NUNC-BD) at the concentration of $8x10^6$ cells/ml for 2h at 37°C 5% v/v CO_2 in serum free medium (Dulbecco's Modified Eagle Medium, DMEM). Non-adherent cells were removed and wells washed twice with Phosphate Buffer Saline (PBS) and 2 ml of DMEM 10% v/v Fetal Bovine Serum, 5% v/v Human AB serum and 1% w/v penicillin/streptomycin were added to each well; 50% of the medium (1.0 ml) was replaced with fresh medium every 3 days. After 7 days of differentiation MDM were treated for 24h with MDA-MB-231 and MDA-MB-468 cancer cell derived supernatant (CM) as previously reported (Kitamura et al., 2015). After 24h all the supernatant was removed and used for quantitative real-time (qPCR) metastasis breast cancer array (see below), cells were washed twice with PBS and lysed with Trizol Reagent (Thermo Fisher) for RNA extraction; RNA was extracted using Trizol manufacturer's protocol. RNA was converted to cDNA using Invitrogen Superscript Vilo cDNA synthesis kit and qPCR was performed using the protocol described above in the text.

iPSC derived macrophages

The SFCi55 iPSC line was generated in house and was confirmed to be pluripotent and have a normal karyotype (C.-T. Yang et al., 2017). The cells were maintained in StemPro medium prepared by supplementing DMEM/F12 + Glutamax (Invitrogen) with StemPro hESC supplement (Invitrogen), 1.8% BSA (Invitrogen), 0.1 mM β-mercaptoethanol (Invitrogen) and 20 ng/ml human basic FGF (Invitrogen). Differentiation of iPSCs to macrophages was carried out as previously described (Lopez-Yrigoyen et al., 2018). When iPSC colonies covered approximately 80% of the culture surface, (Day 0), spent medium was removed and replaced with 1.0 ml StemPro supplemented with cytokine Mix 1 (50 ng/ml BMP4, 50 ng/ml VEGF, and 20 ng/ml SCF). Colonies were cut using the EZPassage™ tool, and gently dislodged with a Pasteur pipette. They were divided equally into two wells of an Ultra-Low Attachment 6-well plate (Corning), and 2 ml of fresh StemPro media with cytokine Mix 1. Cells were cultured in suspension until day 4 with a cytokine top up on Day 2, to make embryoid bodies (EBs). On Day 4, EBs were lifted and transferred to gelatin-coated tissue-culture grade 6-well plates in X-VIVOTM 15 media (Lonza) supplemented with cytokine Mix 2 (100 ng/ml CSF1, 25 ng/ml IL3, 2.0 mM Glutamax, 1% Penicillin/Streptomycin, 0.055 M β-mercaptoethanol). 10 to 15 EBs were plated in each well. EBs were maintained in this medium for the remainder of the protocol, with spent medium being replaced with fresh medium every 3-4 days. After about 2-3 weeks, the EBs produced macrophage progenitors in the culture supernatant that were harvested and transferred to 10 cm^2 bacteriological dishes in X-VIVOTM 15 medium supplemented with cytokine Mix 3 (100 ng/ml CSF1, 2.0 mM Glutamax, 1% v/v Penicillin/Streptomycin) and

allowed to mature for 7 days into iPSC-derived macrophages (iPSC-DM). Macrophage progenitors were harvested every 4 days for approximately 2 months.

THP-1 monocyte differentiation and cytokine stimulation

Human THP-1 monocytes were maintained in culture medium (10% v/v Fetal Bovine Serum [FBS] Roswell Park Memorial Institute [RPMI] 1640 Medium) and incubated at 37°C in a 5% v/v CO_2 atmosphere. For monocyte-macrophage differentiation, cells were seeded in at a density of $2.5x10^5$ cells/ml on 12-well plates, or $5x10^5$ cells/ml in 6-well plates and macrophage differentiation was initiated by exposing the cells to 5ng/ml phorbol-12-myristate-13-acetate (PMA) (Sigma-Aldrich, 16561-29-8) in 10% v/v FBS culture medium at 37°C in a 5% v/v CO_2 atmosphere for 24h. Subsequently, THP-1 derived macrophages were polarized using different combinations of IL-4, IL-10, IL-13 and TGF-β (InvivoGen, USA) or using different pro-inflammatory cytokines including TNFα, IFNγ IL-1β, IL-6 and IL-12 (InvivoGen, USA). The cytokines doses were 20 ng/ml and LPS was used at 25 ng/ml.

Cancer cell culture, conditioned medium production and cytokine stimulation

MDA-MB-468 and THP1 cell lines were cultured in (RPMI1640 with 10% v/v serum (GIBCO, Life Technologies); MDA-MB-231 cells were cultured in DMEM with 10% v/v serum (GIBCO, Life Technologies). All cells were originally obtained from ATCC (Manassas, VA, USA) and subsequently maintained in our laboratory. All cell lines were frequently tested for mycoplasma contamination using a commercially available Mycoplasma detection kit (Myco alert kit, Lonza, USA), and all tested negative. To obtain MDA-MB-231 and -468 CMs cells were resuspended in culture medium, seeded at a density of $1x10^5$ cells/ml in 2.5 ml culture medium on 6-well plates and cultured overnight at 37°C in a 5% v/v CO_2 atmosphere. Subsequently, for CM exposure on PMA-THP-1 monocytes, culture medium was replaced with 10% v/v FBS RPM1640 medium, for CM exposure on human monocyte-derived macrophages (MDMs), culture medium was replaced with 10% v/v FBS DMEM supplemented with 5% v/v human serum and for CM exposure on human iPSDM culture medium was replaced with 10% v/v FBS DMEM. After medium change, cells were cultured for an additional 24h with fresh medium and thereafter, cell free supernatants were harvested and directly used for the experiment.

CCL8 mRNA, SIGLEC1 and CD163 protein detection

Staining was performed on a Leica RX research staining robot (Milton Keynes UK). RNAscope (ACD Bio Newark, CA) In situ hybridization was performed using manufacturer's recommendations

using ACD 2.5 Brown kit (322100). mRNA integrity was assessed using PPIB (cat 313908) on FFPE fixed Breast CA needle biopsies. Mild conditions (ER2 95 C 5mins) with protease (5 mins) were assessed as optimal tissue pre-treatments. Modification of standard protocol for CCL8 (466498) involved omitting DAB substrate and replacing with Tyramide Cy5 (Perkin Elmer, Waltham, Massachusetts). Sections were then sequentially stained for SIGLEC1 protein with Tyramide Cy3 detection and CD163 with Tyramide FITC detection using heat elution between detections for specificity as described (Tóth and Mezey, 2007).

ELISAs

Human CCL8, TNFα, and IL1β protein levels were quantified by Duoset ELISA kits (RandD systems, USA) following manufacturer's instructions. Human CSF1 protein levels were quantified by quantikine ELISA kit (RandD systems, USA). All cell culture supernatants were used undiluted. CCL2 levels were assessed in plasma from 15 healthy donors and 42 breast cancer patients using Legendplex bead-based immunoassays (Biolegend) according to manufacturer's protocol. Data were collected using the C4 Accuri (BD). ELISA for human CX3CL1 was done using a human CX3CL1 Quantikine ELISA kit (R&D Systems) as per manufacturer's instructions.

Cytokine array

Human Cytokine ELISA Plate Array (Signosis, EA-4002), consisting of one pre-coated plate able to detect 32 cytokines simultaneously for 3 independent human samples was used to quantify cytokines in supernatants from MDMs before or after cancer CM stimulation. Detection of cytokines produced from MDMs before or after CM stimulation was performed based on the manufacturers instructions) (Sigma-Aldrich, 16561-29). 8.0 µl of MDM supernatants from each group was added into each well of the plate and incubated at room temperature for 2h. After washing, 100µl of diluted biotin-labeled antibody mixtures were added into each well for another one hour incubation. After washing again, each well was incubated with detection antibody mix and then HRP, and the plate was read on a plate reader at 450 nm. The full list of proteins detected and raw data can be found in **File S4A**.

iPSDM-Cancer cell conditioned medium production

Human iPSDM culture medium was replaced with 10% v/v FBS DMEM 24h before CM incubation. iPSDM were incubated with MDA-MB-231 and -468 CMs (prepared as described above) for 24h and then medium was changed; after medium change, cells were cultured for an additional 24h with fresh medium and thereafter, cell free supernatants were harvested and directly used for the experiment.

TNFα neutralization in iPSDM-Cancer cell conditioned medium

iPSDM-Cancer cell conditioned medium was incubated for 24h with 1.0 µg/ml of mouse anti human TNFα neutralizing antibody (RandDsystems, USA, MAB210-SP, Clone 1825) or 1.0 µg/ml ofmouse IgG_1 isotype control (RandDsystems, USA, MAB002). Efficacy of anti-TNFα antibody neutralization was tested by TNFα ELISA before use.

PCR Arrays

PCR-based microarrays for evaluating the expression of genes mediating the inflammatory response were performed using the human inflammatory cytokines and receptors RT2 Profiler TM PCR array (Qiagen, PAHS-011ZE-4); PCR-based microarrays for evaluating the expression of genes in breast cancer cell lines were performed using the Breast cancer PCR array RT2 Profiler TM PCR array (Qiagen, PAHS-131Z-4) and the Tumor Metastasis PCR array RT2 Profiler TM PCR array (Qiagen, PAHS-028Z). The arrays were configured in a 384-well plate consisted of a panel of 92 genes and 4 endogenous genes. Reverse transcription was performed using the RTC First Strand Kit (Qiagen, 330401) and qPCR was performed using RTC SYBR Green/ROX PCR Master mix (Qiagen, 330521), and the raw data were analysed by the GeneGlobe Data Analysis Center (www.qiagen.com) according to the manufacturer's instructions.

Cell proliferation assay

Cell proliferation was determined using the Cell Counting Kit (CCK)-8 assay (Sigma-Aldrich, 96992) according to the manufacturer's instructions. A total of 5,000 cells were seeded into each well in the 96-well plates and allowed to attach overnight. Cells were then treated with 0.1 ng/ml, 1.0 ng/ml or 10 ng/ml CCL8 (R&D Systems, 281-CP-010/CF). After the treatment (6 to 72 hours), a CCL8 solution was added to each well and then cells were incubated at 37°C for 2h. Cell proliferation was measured using the microplate reader and the proliferation of cells was defined as OD450-OD620.

In vitro cell migration assay

The "scratch" assay to assess cell migration was performed following previously published protocols (Liang et al., 2007). Cells were grown in DMEM with 10% v/v FBS in 12-well plates until they reached confluence; after 24h of starvation (DMEM 0% FBS), a scratch was performed using a p200 eppendorf tip. Recombinant CCL8 and CCL2 were used at 1ng/ml concentrations in all the experiments. Cells were filmed for 24h in a 37°C thermostatic chamber using an Axiovert Scope. 2-3 independent sections/well were filmed and 4 independent experiments per condition were performed.Data analysis was performed with Image J (NIH).

Chemotaxis assay

THP1 chemotaxis was performed using Essen Biosciences reagents. THP1 cells were cultivated in RPMI medium with 10% FBS and seeded at 4000 cells/well in 96 well chemotaxis plates in the presence or absence of 20 ng/ml recombinant human CCL2 or CCL8 (R and D systems). Migration was recorded every hour for 72h using the IncuCyte system (Essen Bioscience) and number of cells migrated was calculated using IncuCyte quantification software.

Statistics

Statistical significance was calculated by Student's t-test when comparing two groups or by one-way or two-way ANOVA when comparing three or more groups. A p-value < 0.05 was considered as statistically significant.

Accession numbers

The transcriptome data reported in the paper have been deposited in the GEO database under accession numbers GSE100925 and GSE117970 (token to access data uvglgwuynvknxod).

Main Figure Legends

Figure 1. Cancer alters the transcriptome of human monocytes.

A) Principal component analysis (PCA) plot of N = 12,157 genes expressed in monocytes from healthy individuals (Mo, N = 45) and TEMo from cancer patients (N = 35; Breast cancer = 32; Endometrial cancer = 3). **B)** Hierarchical clustering of all differentially expressed genes (DEGs) between Mo and TEMo. Expression values are Z score-transformed. Samples were clustered using complete linkage and Euclidean distance. **C)** Gene ontology analysis of DEGs between TEMo and Mo (blue = down-regulated genes, red for upregulated genes). **D)** Bar plot of selected DEGs in TEMo (FDR <= 0.05). **E)** Expression of CD200R1, TNFSF10, HGF and ANGPT1 mRNA in Mo and Breast TEMo (N = 3-5; independent from the RNA-seq cohort). Data are depicted as fold change versus control (Mean±SEM). **F)** Relative distribution of non-classical monocytes from healthy controls and breast and endometrial cancer patients determined by flow cytometry shown as percentage in the monocyte gate. Cohort 1: USA (Mo, N = 31, breast TEMo , N = 22, endometrial TEMo , N = 12); Cohort 2: UK Breast cancer and controls only (Mo, N = 18, TEMo, N = 33). **G)** Quantification of CX3CL1 and CCL2 in the sera of control (Ctr, N = 15) and breast cancer patients (Ca, N = 45). **H)** Differential expression of CX3CR1 and CCR2 in Mo (N = 10) and breast TEMo (N = 31). **I)** Confusion matrix and K) Table showing results of Recursive Feature Elimination with Random Forest (RFE-RF) classification on the testing set (N = 22) for breast TEMo. **L)** Receiver Operating Characteristic (ROC) curves of RFE-RF classification in the training and test set (green and blue, respectively). Monocytes were identified as in **Figure S1** and then stained with CX3CR1 and CCR2 Abs. Data are depicted as net GEO Mean. *p < 0.01, ***p < 0.0001; (E-H) Student's t-test. **See also Figure S1 and File S1.**

Figure 2. TAMs from breast and endometrial cancers exhibit cancer-specific transcriptional profiles.

A) PCA plot of N = 13,668 genes expressed in breast tissue resident macrophages (Br-RM, N = 4) and breast cancer TAMs (Br-TAM, N = 4). **B)** Hierarchical clustering of all DEGs between Br-RM and Br-TAM. Expression values are Z score-transformed. Samples were clustered using complete linkage and Euclidean distance. **C)** Gene ontology analysis of DEGs between Br-TAM and Br-RM (blue = down-regulated genes, red for upregulated genes). **D)** Bar plot of selected DEGs in Br-TAM (FDR <= 0.05). **E)** Venn diagram of commonly regulated transcripts in Br-TAM and TEMo (Red: up-regulation; Blue: down-regulation)**. F)** PCA plot of N = 13,739 genes expressed in endometrial tissue resident macrophages (En-RM, N = 5) from healthy individuals and endometrial cancer TAMs (En-TAM, N = 9). **G)** Hierarchical clustering of all DEGs between En-RM and En-TAM. Expression values are Z score-transformed. Samples were clustered using complete linkage and Euclidean distance. **H)** Gene ontology analysis of DEGs between En-TAM and En-RM (blue = down-regulated

genes, red for upregulated genes).. **I)** Bar plot of selected DEGs in En-TAM (FDR <= 0.05). **J)** Venn diagram of commonly regulated transcripts between En-TAM and TEMo and **K)** En-TAM and Br-TAM. (Red: up-regulation; Blue: down-regulation). **L)** TAM signature score stratified by the CSF1 signature in the DUKE cohort (N = 47) and **M)** by PAM50 molecular subtypes in the METABRIC cohort (N = 1350). (Adjusted p-values: one-way ANOVA with Tukey's post-hoc multiple comparisons test). **N)** Disease-specific survival of the TAM signature expression on the METABRIC cohort (N = 1350). ***p < 0.0001. **See also S2, S3 and File S1.**

Figure 3. Breast TAM transcriptomes are associated with clinical outcomes and reveal new TAM-specific markers.

A) Scatterplot showing Pearson's correlation between CD163 and SIGLEC1 gene expression in the METABRIC cohort (N =1350). Red line indicates local regression (LOESS) fit. **B)** Disease-specific survival of SIGLEC1 gene expression on the METABRIC cohort (N = 1350). **C)** Expression of SIGLEC1 mRNA in BR-RM (N = 4) and Br-TAM (N = 6). **D)** SIGLEC1 expression on public datasets. (Left) Finak et al. (Finak et al., 2008) (p = 0.0002; Normal stroma, N = 6; Cancer stroma, N = 53). (Right) Karnoub et al. (Karnoub et al., 2007) (p = 2.3e-05; Normal stroma, N = 15; Cancer stroma, N = 7). SIGLEC1 expression was calculated from the median centered normalised values. P-values were estimated using a Wilcoxon rank sum test. **E)** CD163 and SIGLEC1 immunofluorescent (IF) staining. (Left to right) SIGLEC1, DAPI, CD163, merged stains and inset (N=5). Independent stains from two different cancer samples (i, ii) are shown representative of N = 12 independent tumors analyzed. Scale bars = 50 μ m and inset = 5 μ m. **F)** Quantification of CD163 (left panel), SIGLEC1 (center panel), and CD163 and SIGLEC1 (right panel) positive cells per mm^2 of tissue in benign (N = 4) and breast cancer samples (N = 8) **G)** TNF α levels in supernatants of iPSDM incubated for 24h with PBS, PBS plus isotype control and PBS plus anti-TNF α antibody (grey bars). Same conditions are shown for MDA-MB-231 (pink bars) and MDA-MB-468 CM (blue bars) (N = 3). Results are expressed as pg/ml. **H)** Expression of TNF α mRNA in Br-RM (N = 4) and Br-TAM (N = 6). **I)** TNF α protein levels in MDM supernatants incubated for 24h with MDA-MB-231 and MDA-MB-468 conditioned medium (CM). Results are expressed as OD at 450nm (N = 3). The full list of proteins detected by the protein array can be found in **File S4D. J)** SIGLEC1 mRNA expression in iPSDM stimulated for 24h with MDA-MB-231 CM normalized as 1 (pink bar), MDA-MB-231 CM + TNF α neutralizing antibody (green bar) and MDA-MB-231 CM + isotype control antibody (magenta bar) (N = 3). **K)** SIGLEC1 mRNA expression in iPSDM stimulated for 24h with MDA-MB-468 CM normalized as 1 (blue bar), MDA-MB-468 CM + TNF α neutralizing antibody (green bar) and MDA-MB-468 CM + isotype control antibody (magenta bar). (N = 3). *p < 0.01, **p <0.001, ***p < 0.0001, ****p < 0.00001; (C, H) Student's t-test, (F) Two-way ANOVA, (G, I-K) One-way ANOVA, (C, F-

K) Data on expression are presented as Mean±SEM, (C, H, J, K) Data are depicted as fold change vs control. **See also S4, S5, S6 and File S4D.**

Figure 4. TAMs and cancer cells engage in cytokine feedback loops to support CCL8 and SIGLEC1 expression in breast cancer TAMs.

A) Volcano plot showing genes whose expression was significantly (Log_2FC +/- 2, $p < 0.05$) deregulated in PMA treated THP1 cells after incubation with MDA-MB-231 for 24h (N = 3), and **B)** after incubation with MDA-MB-468 for 24 h (N = 3). **C)** Venn diagram of commonly up-regulated transcripts between MDA-MB-231 (left circle) and MDA-MB-468 (right circle) treated THP1 cells. **D)** Selection of pro-inflammatory genes commonly up-regulated in Br-TAM (N = 4) (from RNA-seq analysis) and PMA-THP1 (N = 3) (qPCR). **E)** Scatterplot showing Pearson's correlation between CD163 and CCL8 gene expression in the METABRIC cohort (N =1350). Red line indicates local regression (LOESS) fit. **F)** Disease-specific survival of CCL8 expression on the METABRIC cohort (N = 1350). **G)** CCL8 mRNA expression in Br-RM (N = 4) and Br-TAM (N = 7). **H)** CCL8 levels in CM from MDA-MB-231, MDA-MB-468, MDM, and MDM incubated for 24h with the two cancer cell CM respectively. Data are depicted as pg/ml (N = 3). **I)** FISH for CCL8 mRNA. Samples were processed using IF and FISH for (i) CCL8 or (ii) a DapB negative control probe. Depicted are 100x maximum-projected confocal as follows (left to right): SIGLEC1 (red), DAPI (blue), CD163 (green), CCL8 mRNA (white), overlay, magnified inset, and magnified orthogonal planes. All scale bars are 10m. Upper panel: stained cancer sample, lower panel: negative control (N = 3). CCL8 mRNA expression in iPSDM stimulated for 24h with **J)** MDA-MB-231 CM normalized as 1, MDA-MB-231 CM + TNFα neutralizing antibody and MDA-MB-231 CM + isotype control antibody, **K)** MDA-MB-468 CM normalized as 1, MDA-MB-468 CM + TNFα neutralizing and MDA-MB-468 CM + isotype control antibody (N = 3, each). **L)** CSF1 levels and **M)** TNFα and IL1β levels in supernatants from unstimulated MDA-MB-231 (CTR), and MDA-MB-231 incubated for 24h with 10 or 20ng/ml (or 20ng/ml for CSF1) of rCCL8. Unstimulated MDA-MB-468 (CTR), and MDA-MB-468 incubated for 24h with 10 or 20ng/ml (or 20ng/ml for CSF1) of rCCL8. Data are depicted as pg/ml (N = 3 for each). **N)** *In vitro* scratch assay on MDA-MB-231, (left) on untreated CCL8, and (right) CCL2 treated cells (upper panel 1h, lower panel 24h time points, red line = cell culture margins N = 4). **O)** Quantification of *in vitro* scratch assay covered by MDA-MB-231 after 24h (calculated as area covered at 24h –1h) in untreated (CTR), CCL8- and CCL2- treated cells. Same symbols represent Mean of technical replicates (N = 4). **P)** THP1 chemotaxis assay for CCL2 and CCL8. Cells were incubated with medium alone (CTR) or with 20ng/ml of rhCCL2 or rhCCL8. Results shown as fold change vs CTR at 72h (N = 3). *$p < 0.01$, **$p < 0.001$, ***$p < 0.0001$; (G) Student's t-test, (H, J-M, O) One-way ANOVA, (G, H, J, K, M, O) Data on expression are presented as Mean±SEM, (G, J, K) Data are depicted as fold change vs control. **See also S7, S8 and S9.**

Figure 5. High expression of SIGLEC1/CCL8 is associated with poor outcomes in breast cancer patients. Heatmap and disease-specific survival of SIGLEC1 and CCL8 gene expression on **A)** the breast cancer stroma dataset (Finak et al., 2008), **B)** the METABRIC cohort and **C)** the ER+/HER2- patients from the METABRIC cohort. All significant cut-points ($p < 0.05$) are shown in black. Black vertical lines indicate positivity for ER and HER2 expression or grade III tumours.

Figure 6. Schematic representation of the crosstalk between Br-TAM and cancer cells. Tumor cells up-regulate SIGLEC1, TNFα and CCL8 expression in Br-TAM. In turn cancer cells respond to CCL8 stimulation by producing CSF-1, IL1β and TNFα, which further contribute to the positive feedback loop.

Supplementary Figure Legends

Figure S1, related to Figure 1. Flow cytometry gating strategy for the identification and isolation of human monocytes and transcriptomic analysis of the non-classical sub-population.

A) Representative monocyte gating strategy based on physical and fluorescence parameters Validation of gating strategy was performed by **B)** backgating and nuclei coloration (Giemsa, staining).

C) Representative monocyte gating strategy based on physical, and fluorescence parameters Classical and non-classical monocytes separation in **D)** healthy controls and **E)** cancer patients.

F) PCA plot of N = 12,712 genes expressed in non-classical monocytes derived from healthy individuals (N = 6), breast cancer (N = 6) and endometrial cancer patients (N = 7).

G) Hierarchical clustering on all samples (Healthy, N=5; Breast cancer, N=6; Endometrial cancer, N = 7) using the significantly DEGs between breast and healthy non-classical monocytes (N = 139 genes). Expression values are Z score-transformed. Samples were clustered using complete linkage and Euclidean distance.

H) Plot of the accuracy yielded for different gene sizes during feature selection with Recursive feature elimination with Random forest (RFE-RF) model training using a 5 times 10-fold cross validation. X-axis shows the number of genes and y-axis their accuracy. 17 genes showed the best performance.

I) Histogram showing the performance of random classifiers during Random forest (RF) model training. Random classifiers were tested using a 5 times 10-fold cross validation. Solid vertical red line represents the performance of the observed 17-gene signature on the training data.

Figure S2, related to Figure 2. Flow cytometry sorting strategy of tissue macrophages and TAMs, and analysis of CD163 expression and distinct macrophage populations.

A) Gating strategy for tissue macrophages and TAMs; macrophages were defined as CD45+CD3/56/19-CD11b+CD14+CD163+.

B) Representative histogram of macrophage CD163 expression; black = CD163 staining, light grey = Fluorescence minus one (FMO) control. CD163 expression in Br-MR (N = 5) compared to Br-TAM (N = 5). Data are expressed as Geometric Mean (Mean±SEM).

C) (Left) PCA plot of N = 14,229 expressed genes in Br-TAM (N = 4) and En-TAM (N = 9). (Right) Hierarchical clustering of all DEGs between Br-TAM and En-TAM. Expression values are Z score-transformed. Samples were clustered using complete linkage and Euclidean distance.

D) (Left) PCA plot of N = 13,907 expressed genes in Br-RM (N = 4) and En-RM (N = 5). (Right) Hierarchical clustering of all DEGs between Br-RM against En-RM. Expression values are Z score-transformed. Samples were clustered using complete linkage and Euclidean distance.

E) (Top to Bottom) Enrichment analysis of M1-like macrophage signature in Br-TAM: Plot suggests that M1-like genes (Martinez et al., 2006) to be positively (NES = 2.16) and significantly correlated (FDR = 0.003) (left peak) with Br-TAM expression. Black bars represent the position of M1-like genes in the ranked list of Br-TAM expressed genes together with the running enrichment score (green line). Enrichment analysis of M2-like macrophage signature in Br-TAM: Plot suggests that M2-like genes to be positively (NES = 2.22) and significantly correlated (FDR = 0.003) (left peak) with Br-TAM expression. Black bars represent the position of M2-like genes in the ranked list of Br-TAM expressed genes together with the running enrichment score (green line).

F) (Left to right) Enrichment analysis of M1-like macrophage signature in En-TAM: Plot suggests that M1-like genes (Martinez et al., 2006) to be negatively (NES = -1.63) but not significantly correlated (FDR = 0.10) (right peak) with En-TAM expression. Black bars represent the position of M1-like genes in the ranked list of En-TAM expressed genes together with the running enrichment score (green line). Enrichment analysis of M2-like macrophage signature on En-TAM: Plot suggests that M2-like genes to be to be negatively (NES = -1.32) but not significantly correlated (FDR = 0.17) with En-TAM expression. Black bars represent the position of M2-like genes in the ranked list of En-TAM expressed genes together with the running enrichment score (green line).

Figure S3, related to Figure 2. Breast cancer TAM signature is associated with aggressive breast cancer molecular subtypes in the DUKE cohort.

A) TAM signature score across breast cancer subtypes in the DUKE cohort. TAM signature score is calculated as the median of expression of the TAM signature (N = 37) using the median centered normalised FPKM values. Adjusted p-values were estimated using one-way ANOVA with Tukey's post-hoc multiple comparisons test.

B) Hierarchical clustering of the TAM signature on the DUKE cohort tumor samples (N = 47). Expression values of TAM signature were median centered; genes were clustered using average distance and Pearson correlation.

C) The previously identified macrophage immune signature (Bindea et al., 2013; Tamborero et al., 2018) stratified by PAM50 molecular subtypes in the METABRIC cohort. The macrophage signature score was calculated as the median of expression of the macrophage immune signature (N = 29 out of 33 genes were found) using the median centered normalised values. Adjusted p-values were estimated using one-way ANOVA with Tukey's post-hoc multiple comparisons test.
D) Disease-specific survuval of macrophage immune signature expression on the METABRIC cohort (N = 1350). P-values are based on Wald test.
*p < 0.05, ***p < 0.00001.

Figure S4, related to Figure 3. Expression of SIGLEC1 in Br-TAM and other immune cell types.
A) SIGLEC1 expression in 4 independent biological replicate Br-TAM samples (macrophages gated as in **Figure S2**); (upper panel) FMO controls, (lower panel) SIGLEC1 staining.
B) GEO Mean of SIGLEC1 expression in Br-TAM (white bar) and in FMO controls (black bar).
C) SIGLEC1 expression in CD3/19/56 positive cells from breast tumors; (upper panel) FMO controls, (lower panel) SIGLEC1 staining.
D) GEO Mean of SIGLEC1 expression in CD3/19/56 positive cells from breast tumors and in FMO controls.
E) SIGLEC1 expression in 4 independent breast tumor samples (CD45-): (upper panel) FMO controls, (lower panel) SIGLEC1 staining.
F) GEO Mean of SIGLEC1 expression in CD45- cells from breast tumors and in FMO controls.
G) Quantification of % of SIGLEC1 positive cells in Br-TAM (N = 3). CD3/19/56 positive cells and CD45- cells from breast tumors. Data are depicted as % of SIGLEC1 positive cells (Mean±SEM),
H) GEO Mean values of SIGLEC1 expression in Br-TAM (N = 4). CD3/19/56 positive cells and CD45- cells from breast tumors. Data are depicted as GEO Mean (Mean±SEM)
I) Representative SIGLEC1 cell surface expression in classical and non-classical Mo and breast cancer TEMo (monocytes gated as in **Figure S1**), square gate based on FMO control.
K) Quantification of % of SIGLEC1 positive cells in Mo and TEM (N = 3). Data are depicted as % of SIGLEC1 positive cells (Mean±SEM).
L) Representative SIGLEC1 cell surface expression in breast cancer granulocytes (gated on live CD45+ SSC high cells,), square gate based on FMO control.
M) Quantification of % of SIGLEC1 positive cells in granulocytes from normal patients (CTR) and breast cancer patients (N = 3, Mean±SEM).
*p < 0.01, **p < 0.001; (G,H) One-way ANOVA, (K,M) Student's t-test.

Figure S5, related to Figure 3. Flow diagram showing the acquisition and analysis pipeline for cell quantification of immunofluorescently stained breast cancer tissues.

A) Images of 40x-scanned slides were processed in Tissue Studio for nuclear detection and cell simulation. Nuclear/cell quality control and manual sampling of cell classes were conducted in Developer XD and a 19-dimensional feature space optimized from 49 metrics and ranked by importance to Euclidean distance separation. Initial classification of tissues was refined 10 or more times using cycles of classification-correction-reclassification. This process was repeated for additional channels, then entire batches of scanned slides processed together and cell classification data exported. Final extraction, compilation, and statistics of cell numbers were conducted using Mathematica and GraphPad Prism as described in materials and methods.
B) Additional CD163 and SIGLEC1 immunofluorescent staining on breast cancer tissue samples. Tissue biopsies were stained for SIGLEC1, CD163 and DAPI. (Left to right) SIGLEC1, DAPI, CD163, merged stains and inset; 60X magnification. Breast cancer and benign (N = 5 each, Bars = 50 μm). Shown 5 cancers and one benign control.

Figure S6. Multiplex immunohistochemistry analysis of different breast cancer subtypes.
A) Tissue biopsies from 13 breast cancer patients and three prophylactic mastectomy/mammoplasty (PM) procedures were assessed using quantitative multiplex immunohistochemistry. Cumulative cell populations (indicted on right) from total tissue areas were normalized to total cell number.
B) Representative micrographs reflecting pseudo-colored images following multiplex IHC of cell populations across breast cancer subtypes as indicated. Boxed insets are depicted at higher magnification in corresponding columns. Scale bars as indicated.
C) A heatmap of each cell population as a percent of total cells is shown with a dendrogram of unsupervised hierarchical clustering, scaled by row and using correlation as a distance measure, and average as a clustering method. Each column represents an independent tumor according to sub-type. (Lum: luminal) and prophylactic mastectomy/mammoplasty (PM) samples.

Figure S7, related to Figure 4. Expression of SIGLEC1 and CCL8 in macrophages after stimulation with cancer cell conditioned medium or cytokines.
A) *SIGLEC1* mRNA expression in primary MDM stimulated for 24h with PBS (black bar) normalized as 1, MDA-MB-231 conditioned medium (CM) (red bar) or MDA-MB-468 CM (blue bar). Data are depicted as fold change vs CTR (N = 3, Mean±SEM).
B) *SIGLEC1* mRNA expression in PMA-treated THP1 cells stimulated for 24h with PBS (black bar) normalized as 1, MDA-MB-231 CM (red bar) or MDA-MB-468 CM (blue bar). Data are depicted as fold change vs CTR (N = 3, Mean±SEM).
C) Flow cytometric analysis of SIGLEC1 expression in iPSDM cells stimulated with MDA-MB-231 CM. (Left panel) representative histogram showing SIGLEC1 expression levels in unstained (grey histogram), PBS (black histogram) and MDA-MB-231 CM treated samples (red histogram). (Right

panel) SIGLEC1 expression (Geometric mean expressed as fold change vs CTR) in PBS (black bar) and MDA-MB-231 CM treated samples (red bar). (N = 3, Mean±SEM).

D) Flow cytometric analysis of SIGLEC1 expression in iPSDM cells stimulated with MDA-MB-468 CM. (Left panel) representative histogram showing SIGLEC1 expression levels in unstained (grey histogram), PBS (black histogram) and MDA-MB-468 CM treated samples (blue histogram). (Right panel) SIGLEC1 expression (Geometric mean expressed as fold change vs CTR) in PBS (black bar) and MDA-MB-468 CM treated samples (blue bar). (N = 3, Mean±SEM).

E) *SIGLEC1* mRNA expression in PMA-treated THP1 cells stimulated for times shown with pro-inflammatory cytokines as indicated. LPS acts as a positive pro-inflammatory signal control. Red bar indicates SIGLEC1 expression in treated samples, dotted black line indicates SIGLEC1 expression in control PBS-treated samples. Data are depicted as fold change vs CTR (N = 3).

F) *SIGLEC1* mRNA expression in PMA-treated THP1 cells stimulated for times shown with anti-inflammatory cytokines as indicated. Blue bar indicates SIGLEC1 expression in cytokine-treated samples, dotted black line indicates SIGLEC1 expression in PBS-treated samples. Data are depicted as fold change vs CTR (N = 3).

G) *CCL8* mRNA expression in PMA-treated THP1 cells stimulated for 24h with PBS (CTR; black bar), MDA-MB-231 CM (red bar) or MDA-MB-468 CM (blue bar). Data are depicted as fold change vs CTR (N = 3, Mean±SEM).

H) *CCL8* mRNA expression in primary MDM stimulated for 24h with PBS (CTR; black bar), MDA-MB-231 CM (red bar) or MDA-MB-468 CM (blue bar). Data are depicted as fold change vs CTR (N = 3 Mean±SEM).

I) *CCL8* mRNA expression in iPSDM stimulated for 24h with PBS (CTR; black bar), MDA-MB-231 CM (red bar) or MDA-MB-468 CM (blue bar). Data are depicted as fold change vs CTR (N = 3, Mean±SEM).

J) ELISA for CCL8 in serum of healthy donors (black circles, N = 21) and breast cancer patients (orange squares, N = 38). Data are depicted as pg/ml (Mean±SEM).

K) *CCL8* mRNA expression in PMA-treated THP1 cells stimulated for times shown with pro-inflammatory cytokines as shown. Red line cytokine-treated samples, dotted black line PBS-treated samples; Data are depicted as fold change vs CTR (N = 3, Mean±SEM).

L) *CCL8* mRNA expression in PMA-treated THP1 cells stimulated for times shown with anti-inflammatory cytokines as indicated. Blue bar cytokine-treated samples, dotted black line PBS-treated samples. Data are depicted as fold change vs CTR (N = 3, Mean±SEM).

*p < 0.01, **p < 0.001, ***p<0.0001, ****p<0.00001; (A, B, G-I) One-way ANOVA, (C-F, K, L) Student's t-test.

Figure S8, related to Figure 4. Cancer cell line expression of CCL8 Receptors.

A) Representative histograms of CCR1 (blue), CCR2 (red), CCR3 (green), CCR5 (light blue) and CCR8 (gold) expression in MDA-MB-231 (upper panel) and 468 (lower panel) cells. Grey histograms indicate unstained samples.

B) Percentage of CCR1 (blue), CCR2 (red), CCR3 (green), CCR5 (light blue) and CCR8 (gold) positive cells in total MDA-MB-231 (left panel) and MDA-MB -468 cells (right panel).

C) MDA-MB-231 (left panel) and MDA-MB-468 (right panel) proliferation assay in the presence of PBS (No treatment; black line), 0.1ng/ml (red line), 1ng/ml (blue line) and 10ng/ml (green line) of CCL8 from 0-80h. No statistical differences between treatments and controls (N = 3, Mean±SEM).

Figure S9, related to Figure 4. Breast cancer qPCR array on cancer cells stimulated with rCCL8 and macrophage conditioned medium.

A) Volcano plot showing genes whose expression is significantly (Log_2FC +/- 2, $p < 0.05$) down-regulated (blue dots) and up-regulated (orange dots) in MDA-MB-231 cells after incubation with 1ng/ml of rCCL8 for 16h (N = 3).

B) Volcano plot showing genes whose expression is significantly (Log_2FC +/- 2, $p < 0.05$) down-regulated (blue dots) and up-regulated (orange dots) in MDA-MB-468 cells after incubation with 1ng/ml of rCCL8 for 16h (N = 3).

C) Venn diagram of commonly up-regulated genes between MDA-MB-231 (left circle) and MDA-MB-468 (right circle) after rCCL8 treatment. Expression of 6 genes was commonly up-regulated.

D) mRNA expression of 6 commonly up-regulated genes in MDA-MB-231 or MDA-MB-468 after CCL8 stimulation; dotted black line represents normalized expression level in untreated control samples. Data are depicted as fold change vs CTR (N = 3, Mean±SEM).

E) Volcano plot showing genes whose expression is significantly (Log_2FC +/- 2, $p < 0.05$) down-regulated (blue dots) and up-regulated (orange dots) in MDA-MB-231 cells after incubation with condition medium (CM) primed MDM supernatant for 16h (N = 3).

F) Volcano plot showing genes whose expression is significantly (Log_2FC +/- 2, $p < 0.05$) down-regulated (blue dots) and up-regulated (orange dots) in MDA-MB-468 cells after incubation with CM primed MDM supernatant for 16h (N = 3).

G) Venn diagram of commonly upregulated genes between MDA-MB-231 (left circle) and MDA-MB-468 (right circle). Expression of 16 genes was commonly upregulated in the two conditions.

H) mRNA expression of 16 commonly upregulated genes in MDA-MB-231 (light grey bar) or MDA-MB-468 (dark grey bar) after CM primed MDM supernatant stimulation; dotted black line represents normalized expression level in untreated control samples (N = 3, Mean±SEM).

*p < 0.01, **p < 0.001, ***p<0.0001, *****p<0.000001; (D, H) Student's t-test.

References

Anders, S., Pyl, P.T., Huber, W., 2014. HTSeq A Python framework to work with high-throughput sequencing data. doi:10.1101/002824

Andrews, S., 2012. FastQC: A quality control application for high throughput sequence data. … : http://www. bioinformatics. bbsrc. ac. uk/projects/fastqc.

Arwert, E.N., Harney, A.S., Entenberg, D., Wang, Y., Sahai, E., Pollard, J.W., Condeelis, J.S., 2018. A Unidirectional Transition from Migratory to Perivascular Macrophage Is Required for Tumor Cell Intravasation. CellReports 23, 1239–1248. doi:10.1016/j.celrep.2018.04.007

Asano, K., Takahashi, N., Ushiki, M., Monya, M., Aihara, F., Kuboki, E., Moriyama, S., Iida, M., Kitamura, H., Qiu, C.-H., Watanabe, T., Tanaka, M., 2015. Intestinal CD169(+) macrophages initiate mucosal inflammation by secreting CCL8 that recruits inflammatory monocytes. Nature Communications 6, 7802. doi:10.1038/ncomms8802

Azizi, E., Carr, A.J., Plitas, G., Cornish, A.E., Konopacki, C., Prabhakaran, S., Nainys, J., Wu, K., Kiseliovas, V., Setty, M., Choi, K., Fromme, R.M., Dao, P., McKenney, P.T., Wasti, R.C., Kadaveru, K., Mazutis, L., Rudensky, A.Y., Pe'er, D., 2018. Single-cell Map of Diverse Immune Phenotypes Driven by the Tumor Microenvironment 1–91. doi:10.1101/221994

Barbai, T., Fejős, Z., Puskas, L.G., Tímár, J., Rásó, E., 2015. The importance of microenvironment: the role of CCL8 in metastasis formation of melanoma. Oncotarget 6, 29111–29128. doi:10.18632/oncotarget.5059

Beck, A.H., Espinosa, I., Edris, B., Li, R., Montgomery, K., Zhu, S., Varma, S., Marinelli, R.J., van de Rijn, M., West, R.B., 2009. The Macrophage Colony-Stimulating Factor 1 Response Signature in Breast Carcinoma. Clin Cancer Res 15, 778–787. doi:10.1158/1078-0432.CCR-08-1283

Bharat, A., McQuattie-Pimentel, A.C., Budinger, G.R.S., 2017. Non-classical monocytes in tissue injury and cancer. Oncotarget 8, 106171–106172. doi:10.18632/oncotarget.22584

Bin Fang, W., Jokar, I., Zou, A., Lambert, D., Dendukuri, P., Cheng, N., 2012. CCL2/CCR2 chemokine signaling coordinates survival and motility of breast cancer cells through Smad3 and p42/44MAPK dependent mechanisms. J. Biol. Chem. 287, jbc.M112.365999–36608. doi:10.1074/jbc.M112.365999

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A.C., Angell, H., Fredriksen, T., Lafontaine, L., Berger, A., Bruneval, P., Fridman, W.H., Becker, C., Pagès, F., Speicher, M.R., Trajanoski, Z., Galon, J., 2013. Spatiotemporal Dynamics of Intratumoral Immune Cells Reveal the Immune Landscape in Human Cancer. Immunity 39, 782–795. doi:10.1016/j.immuni.2013.10.003

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic - a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England) 30, 2114–2120. doi:10.1093/bioinformatics/btu170

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. doi:10.1023/A:1010933404324

Cassetta, L., Kajaste-Rudnitski, A., Coradin, T., Saba, E., Chiara, Della, G., Barbagallo, M., Graziano, F., Alfano, M., Cassol, E., Vicenzi, E., Poli, G., 2013. M1 polarization of human monocyte-derived macrophages restricts pre and postintegration steps of HIV-1 replication. AIDS 27, 1847–1856. doi:10.1097/QAD.0b013e328361d059

Cassetta, L., Noy, R., Swierczak, A., Sugano, G., Smith, H., Wiechmann, L., Pollard, J.W., 2016. Isolation of Mouse and Human Tumor-Associated Macrophages, in: Tumor Microenvironment, Advances in Experimental Medicine and Biology. Springer International Publishing, Cham, pp. 211–229. doi:10.1007/978-3-319-26666-4_12

Cassetta, L., Pollard, J.W., 2018. Targeting macrophages: therapeutic approaches in cancer. Nature Reviews Drug Discovery 948, 151. doi:10.1038/nrd.2018.169

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., Hackl, H., Trajanoski, Z., 2017. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. CellReports 18, 248–262. doi:10.1016/j.celrep.2016.12.019

Chávez-Galán, L., Olleros, M.L., Vesin, D., Garcia, I., 2015. Much More than M1 and M2 Macrophages, There are also CD169+ and TCR+ Macrophages. Front. Immunol. 6, 22. doi:10.3389/fimmu.2015.00263

Chittezhath, M., Dhillon, M.K., Lim, J.Y., Laoui, D., Shalova, I.N., Teo, Y.L., Chen, J., Kamaraj, R., Raman, L., Lum, J., Thamboo, T.P., Chiong, E., Zolezzi, F., Yang, H., Van Ginderachter, J.A., Poidinger, M., Wong, A.S.C., Biswas, S.K., 2014. Molecular profiling reveals a tumor-promoting phenotype of monocytes and macrophages in human cancer progression. Immunity 41, 815–829. doi:10.1016/j.immuni.2014.09.014

Curtis, C., Shah, S.P., Chin, S.-F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., Gräf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S., Group, M., Langerød, A., Green, A., Provenzano, E., Wishart, G., Pinder, S., Watson, P., Markowetz, F., Murphy, L., Ellis, I., Purushotham, A., Børresen-Dale, A.-L., Brenton, J.D., Tavaré, S., Caldas, C., Aparicio, S., 2012. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486, 346–352. doi:10.1038/nature10983

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2012. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. doi:10.1093/bioinformatics/bts635

Farmaki, E., Chatzistamou, I., Kaza, V., Kiaris, H., 2016. A CCL8 gradient drives breast cancer cell dissemination. Oncogene 2016 35:49 35, 6309–6318. doi:10.1038/onc.2016.161

Feng, A.-L., Zhu, J.-K., Sun, J.-T., Yang, M.-X., Neckenig, M.R., Wang, X.-W., Shao, Q.-Q., Song, B.-F., Yang, Q.-F., Kong, B.-H., Qu, X., 2011. CD16+ monocytes in breast cancer patients: expanded by monocyte chemoattractant protein-1 and may be useful for early diagnosis. Clinical and experimental immunology 164, 57–65. doi:10.1111/j.1365-2249.2011.04321.x

Finak, G., Bertos, N., Pepin, F., Sadekova, S., Souleimanova, M., Zhao, H., Chen, H., Omeroglu, G., Meterissian, S.,

Omeroglu, A., Hallett, M., Park, M., 2008. Stromal gene expression predicts clinical outcome in breast cancer. Nature medicine 14, 518–527. doi:10.1038/nm1764

Franklin, R.A., Liao, W., Sarkar, A., Kim, M.V., Bivona, M.R., Liu, K., Pamer, E.G., Li, M.O., 2014. The cellular and molecular origin of tumor-associated macrophages. Science 344, 921–925. doi:10.1126/science.1252510

Gentles, A.J., Newman, A.M., Liu, C.L., Bratman, S.V., Feng, W., Kim, D., Nair, V.S., Xu, Y., Khuong, A., Hoang, C.D., Diehn, M., West, R.B., Plevritis, S.K., Alizadeh, A.A., 2015. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nature medicine 21, 938–945. doi:10.1038/nm.3909

Hamm, A., Prenen, H., Van Delm, W., Di Matteo, M., Wenes, M., Delamarre, E., Schmidt, T., Weitz, J., Sarmiento, R., Dezi, A., Gasparini, G., Rothe, F., Schmitz, R., D'Hoore, A., Iserentant, H., Hendlisz, A., Mazzone, M., 2015. Tumour-educated circulating monocytes are powerful candidate biomarkers for diagnosis and disease follow-up of colorectal cancer. Gut 1–12. doi:10.1136/gutjnl-2014-308988

Hanna, R.N., Cekic, C., Sag, D., Tacke, R., Thomas, G.D., Nowyhed, H., Herrley, E., Rasquinha, N., McArdle, S., Wu, R., Peluso, E., Metzger, D., Ichinose, H., Shaked, I., Chodaczek, G., Biswas, S.K., Hedrick, C.C., 2015. Patrolling monocytes control tumor metastasis to the lung. Science 350, 985–990. doi:10.1126/science.aac9407

Johnson, W.E., Li, C., Rabinovic, A., 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8, 118–127. doi:10.1093/biostatistics/kxj037

Jung, K., Heishi, T., Khan, O.F., Kowalski, P.S., Incio, J., Rahbari, N.N., Chung, E., Clark, J.W., Willett, C.G., Luster, A.D., Yun, S.H., Langer, R., Anderson, D.G., Padera, T.P., Jain, R.K., Fukumura, D., 2017. Ly6Clo monocytes drive immunosuppression and confer resistance to anti-VEGFR2 cancer therapy. J. Clin. Invest. 127, 3039–3051. doi:10.1172/JCI93182

Karnoub, A.E., Dash, A.B., Vo, A.P., Sullivan, A., Brooks, M.W., Bell, G.W., Richardson, A.L., Polyak, K., Tubo, R., Weinberg, R.A., 2007. Mesenchymal stem cells within tumour stroma promote breast cancer metastasis. Nature 449, 557–563. doi:10.1038/nature06188

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S.L., 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology 14, R36. doi:10.1186/gb-2013-14-4-r36

Kitamura, T., Qian, B., Soong, D., Cassetta, L., Noy, R., Sugano, G., Kato, Y., Li, J., Pollard, J.W., 2015. CCL2-induced chemokine cascade promotes breast cancer metastasis by enhancing retention of metastasis-associated macrophages. J Exp Med 212, 1043–1059. doi:10.1084/jem.20141836

Knüpfer, H., Preiß, R., 2006. Significance of interleukin-6 (IL-6) in breast cancer (review). Breast Cancer Res Treat 102, 129–135. doi:10.1007/s10549-006-9328-3

Kuhn, M., 2015. caret: Classification and Regression Training. Astrophysics Source Code Library.

Leek, J.T., 2014. svaseq: removing batch effects and other unwanted noise from sequencing data. Nucleic acids research 42, 00–a–0. doi:10.1093/nar/gku864

Li, C., Luo, X., Lin, Y., Tang, X., Ling, L., Wang, L., Jiang, Y., 2015. A Higher Frequency of CD14+ CD169+ Monocytes/Macrophages in Patients with Colorectal Cancer. PloS one 10, e0141817. doi:10.1371/journal.pone.0141817

Liang, C.-C., Park, A.Y., Guan, J.-L., 2007. In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro. Nature protocols 2, 329–333. doi:10.1038/nprot.2007.30

Lopez-Yrigoyen, M., Fidanza, A., Cassetta, L., Axton, R.A., Taylor, A.H., Meseguer-Ripolles, J., Tsakiridis, A., Wilson, V., Hay, D.C., Pollard, J.W., Forrester, L.M., 2018. A human iPSC line capable of differentiating into functional macrophages expressing ZsGreen: a tool for the study and in vivo tracking of therapeutic cells. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 373, 20170219. doi:10.1098/rstb.2017.0219

Makita, N., Hizukuri, Y., Yamashiro, K., Murakawa, M., Hayashi, Y., 2015. IL-10 enhances the phenotype of M2 macrophages induced by IL-4 and confers the ability to increase eosinophil migration. Int Immunol 27, 131–141. doi:10.1093/intimm/dxu090

Martinez, F.O., Gordon, S., Locati, M., Mantovani, A., 2006. Transcriptional profiling of the human monocyte-to-macrophage differentiation and polarization: new molecules and patterns of gene expression. The Journal of Immunology 177, 7303–7311. doi:10.4049/jimmunol.177.10.7303

Masuda, H., Zhang, D., Bartholomeusz, C., Doihara, H., Hortobagyi, G.N., Ueno, N.T., 2012. Role of epidermal growth factor receptor in breast cancer. Breast Cancer Res Treat 136, 331–345. doi:10.1007/s10549-012-2289-9

Neubert, N.J., Schmittnaegel, M., Bordry, N., Nassiri, S., Wald, N., Martignier, C., Tillé, L., Homicsko, K., Damsky, W., Maby-El Hajjami, H., Klaman, I., Danenberg, E., Ioannidou, K., Kandalaft, L., Coukos, G., Hoves, S., Ries, C.H., Fuertes Marraco, S.A., Foukas, P.G., De Palma, M., Speiser, D.E., 2018. T cell-induced CSF1 promotes melanoma resistance to PD1 blockade. Science Translational Medicine 10, eaan3311. doi:10.1126/scitranslmed.aan3311

Neve, R.M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F.L., Fevr, T., Clark, L., Bayani, N., Coppe, J.-P., Tong, F., Speed, T., Spellman, P.T., DeVries, S., Lapuk, A., Wang, N.J., Kuo, W.-L., Stilwell, J.L., Pinkel, D., Albertson, D.G., Waldman, F.M., McCormick, F., Dickson, R.B., Johnson, M.D., Lippman, M., Ethier, S., Gazdar, A., Gray, J.W., 2006. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell 10, 515–527. doi:10.1016/j.ccr.2006.10.008

Parker, J.S., Mullins, M., Cheang, M.C.U., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., Quackenbush, J.F., Stijleman, I.J., Palazzo, J., Marron, J.S., Nobel, A.B., Mardis, E., Nielsen, T.O., Ellis, M.J., Perou, C.M., Bernard, P.S., 2009. Supervised risk predictor of breast cancer based on intrinsic subtypes. J. Clin. Oncol. 27, 1160–1167. doi:10.1200/JCO.2008.18.1370

Pearce, D.A., Nirmal, A.J., Freeman, T.C., Sims, A.H., 2017. Continuous Biomarker Assessment by Exhaustive Survival Analysis. bioRxiv 208660. doi:10.1101/208660

Planet, E., 2013. phenoTest: Tools to test association between gene expression and phenotype in a way that is efficient,

structured, fast and scalable. We also provide tools to do GSEA (Gene set enrichment analysis) and copy number variation. R package version 1.28.0.
. doi:10.18129/B9.bioc.phenoTest

Proost, P., Wuyts, A., van Damme, J., 1996. Human monocyte chemotactic proteins-2 and -3: structural and functional comparison with MCP-1. Journal of Leukocyte Biology 59, 67–74. doi:10.1002/jlb.59.1.67

Qian, B., Li, J., Zhang, H., Kitamura, T., Zhang, J., Campion, L.R., Kaiser, E.A., Snyder, L.A., Pollard, J.W., 2011. CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. Nature 475, 222–225. doi:10.1038/nature10138

Ries, C.H., Cannarile, M.A., Hoves, S., Benz, J., Wartha, K., Runza, V., Rey-Giraud, F., Pradel, L.P., Feuerhake, F., Klaman, I., Jones, T., Jucknischke, U., Scheiblich, S., Kaluza, K., Gorr, I.H., Walz, A., Abiraj, K., Cassier, P.A., Sica, A., Gomez-Roca, C., de Visser, K.E., Italiano, A., Le Tourneau, C., Delord, J.-P., Levitsky, H., Blay, J.-Y., Rüttinger, D., 2014. Targeting Tumor-Associated Macrophages with Anti-CSF-1R Antibody Reveals a Strategy for Cancer Therapy. Cancer Cell 25, 846–859. doi:10.1016/j.ccr.2014.05.016

Ritchie, M.E., Phipson, B., Di Wu, Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research 43, gkv007–e47. doi:10.1093/nar/gkv007

Roberts, A., Trapnell, C., Donaghey, J., Rinn, J.L., Pachter, L., 2011. Improving RNA-Seq expression estimates by correcting for fragment bias. Genome biology 12, R22. doi:10.1186/gb-2011-12-3-r22

Robinson, M.D., Oshlack, A., 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome biology 11, R25. doi:10.1186/gb-2010-11-3-r25

Roomi, M.W., Monterrey, J.C., Kalinovsky, T., Rath, M., Niedzwiecki, A., 2009. Patterns of MMP-2 and MMP-9 expression in human cancer cell lines. Oncology Reports 21, 1323–1333. doi:10.3892/or_00000358

Sing, T., Sander, O., Beerenwinkel, N., Lengauer, T., 2005. ROCR: visualizing classifier performance in R. Bioinformatics 21, 3940–3941. doi:10.1093/bioinformatics/bti623

Takeya, M., Komohara, Y., 2016. Role of tumor-associated macrophages in human malignancies: friend or foe? Pathology International 66, 491–505. doi:10.1111/pin.12440

Tamborero, D., Rubio-Perez, C., Muiños, F., Sabarinathan, R., Piulats, J.M., Muntasell, A., Dienstmann, R., Lopez-Bigas, N., Gonzalez-Perez, A., 2018. A Pan-cancer Landscape of Interactions between Solid Tumors and Infiltrating Immune Cell Populations. Clin Cancer Res. doi:10.1158/1078-0432.CCR-17-3509

Tóth, Z.E., Mezey, E., 2007. Simultaneous visualization of multiple antigens with tyramide signal amplification using antibodies from the same species. J. Histochem. Cytochem. 55, 545–554. doi:10.1369/jhc.6A7134.2007

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., Pachter, L., 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology 28, 511–515. doi:10.1038/nbt.1621

Tsujikawa, T., Kumar, S., Borkar, R.N., Azimi, V., Thibault, G., Chang, Y.H., Balter, A., Kawashima, R., Choe, G., Sauer, D., Rassi, El, E., Clayburgh, D.R., Kulesz-Martin, M.F., Lutz, E.R., Zheng, L., Jaffee, E.M., Leyshock, P., Margolin, A.A., Mori, M., Gray, J.W., Flint, P.W., Coussens, L.M., 2017. Quantitative Multiplex Immunohistochemistry Reveals Myeloid-Inflamed Tumor-Immune Complexity Associated with Poor Prognosis. CellReports 19, 203–217. doi:10.1016/j.celrep.2017.03.037

Wang, X., Spandidos, A., Wang, H., Seed, B., 2012. PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. Nucleic acids research 40, D1144–9. doi:10.1093/nar/gkr1013

Williams, C.B., Yeh, E.S., Soloff, A.C., 2016. Tumor-associated macrophages: unwitting accomplices in breast cancer malignancy. NPJ breast cancer 2, 15025. doi:10.1038/npjbcancer.2015.25

Xue, J., Schmidt, S.V., Sander, J., Draffehn, A., Krebs, W., Quester, I., De Nardo, D., Gohel, T.D., Emde, M., Schmidleithner, L., Ganesan, H., Nino-Castro, A., Mallmann, M.R., Labzin, L., Theis, H., Kraut, M., Beyer, M., Latz, E., Freeman, T.C., Ulas, T., Schultze, J.L., 2014. Transcriptome-Based Network Analysis Reveals a Spectrum Model of Human Macrophage Activation. Immunity 40, 274–288. doi:10.1016/j.immuni.2014.01.006

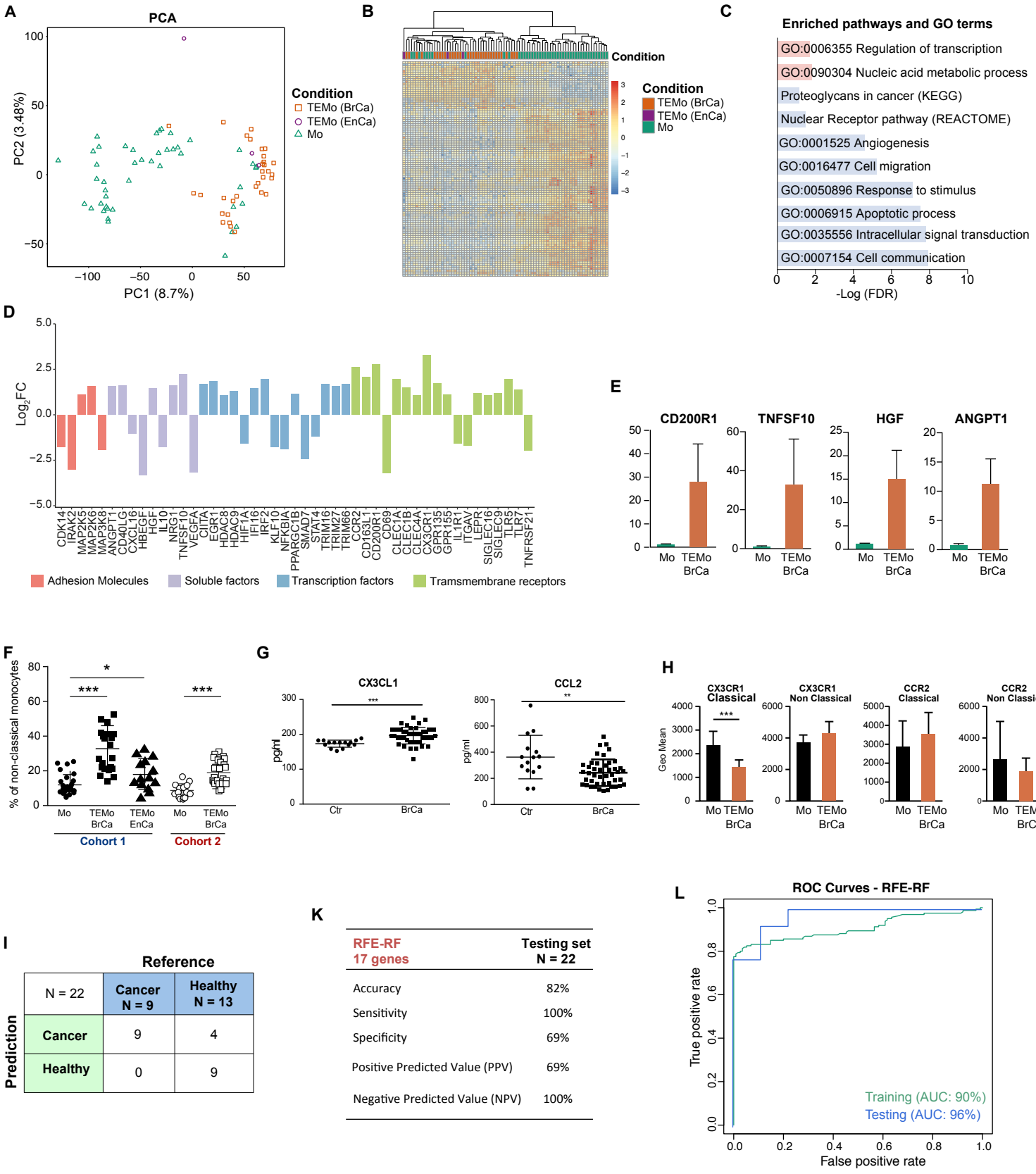
Yang, C.-T., Ma, R., Axton, R.A., Jackson, M., Taylor, A.H., Fidanza, A., Marenah, L., Frayne, J., Mountford, J.C., Forrester, L.M., 2017. Activation of KLF1 Enhances the Differentiation and Maturation of Red Blood Cells from Human Pluripotent Stem Cells. Stem Cells 35, 886–897. doi:10.1002/stem.2562

Yang, M., McKay, D., Pollard, J.W., Lewis, C.E., 2018. Diverse Functions of Macrophages in Different Tumor Microenvironments. Cancer research 78, 5492–5503. doi:10.1158/0008-5472.CAN-18-1367

Zhang, Y., Li, J.-Q., Jiang, Z.-Z., Li, L., Wu, Y., Zheng, L., 2016. CD169 identifies an anti-tumour macrophage subpopulation in human hepatocellular carcinoma. The Journal of Pathology 239, 231–241. doi:10.1002/path.4720

Figure

Figure 1.



I

Prediction	Reference: N = 22	Cancer N = 9	Healthy N = 13
	Cancer	9	4
	Healthy	0	9

K

RFE-RF 17 genes	Testing set N = 22
Accuracy	82%
Sensitivity	100%
Specificity	69%
Positive Predicted Value (PPV)	69%
Negative Predicted Value (NPV)	100%

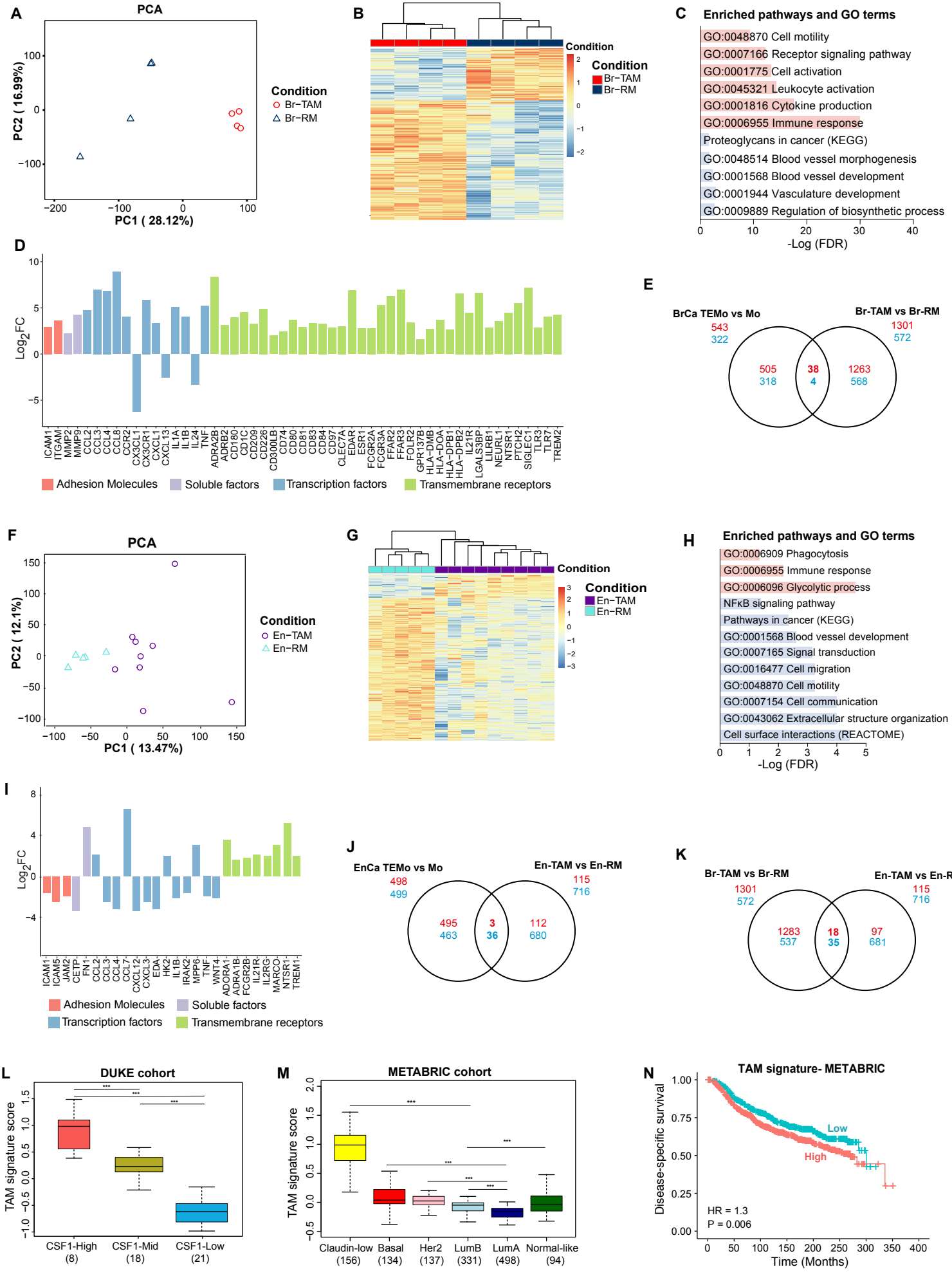

A
PCA
PC2 (16.99%)
PC1 (28.12%)
Condition
Br−TAM
Br−RM
B
Condition
Br−TAM
Br−RM
C
Enriched pathways and GO terms
GO:0048870 Cell motility
GO:0007166 Receptor signaling pathway
GO:0001775 Cell activation
GO:0045321 Leukocyte activation
GO:0001816 Cytokine production
GO:0006955 Immune response
Proteoglycans in cancer (KEGG)
GO:0048514 Blood vessel morphogenesis
GO:0001568 Blood vessel development
GO:0001944 Vasculature development
GO:0009889 Regulation of biosynthetic process
-Log (FDR)
D
Log2FC
Adhesion Molecules
Soluble factors
Transcription factors
Transmembrane receptors
E
BrCa TEMo vs Mo
543
322
Br-TAM vs Br-RM
1301
572
505
318
38
4
1263
568
F
PCA
PC2 (12.1%)
PC1 (13.47%)
Condition
En−TAM
En−RM
G
Condition
En−TAM
En−RM
H
Enriched pathways and GO terms
GO:0006909 Phagocytosis
GO:0006955 Immune response
GO:0006096 Glycolytic process
NFκB signaling pathway
Pathways in cancer (KEGG)
GO:0001568 Blood vessel development
GO:0007165 Signal transduction
GO:0016477 Cell migration
GO:0048870 Cell motility
GO:0007154 Cell communication
GO:0043062 Extracellular structure organization
Cell surface interactions (REACTOME)
-Log (FDR)
I
Log2FC
Adhesion Molecules
Soluble factors
Transcription factors
Transmembrane receptors
J
EnCa TEMo vs Mo
498
499
En-TAM vs En-RM
115
716
495
463
3
36
112
680
K
Br-TAM vs Br-RM
1301
572
En-TAM vs En-RM
115
716
1283
537
18
35
97
681
L
DUKE cohort
TAM signature score
CSF1-High (8)
CSF1-Mid (18)
CSF1-Low (21)
M
METABRIC cohort
TAM signature score
Claudin-low (156)
Basal (134)
Her2 (137)
LumB (331)
LumA (498)
Normal-like (94)
N
TAM signature- METABRIC
Disease-specific survival
Low
High
HR = 1.3
P = 0.006
Time (Months)

Figure 3.

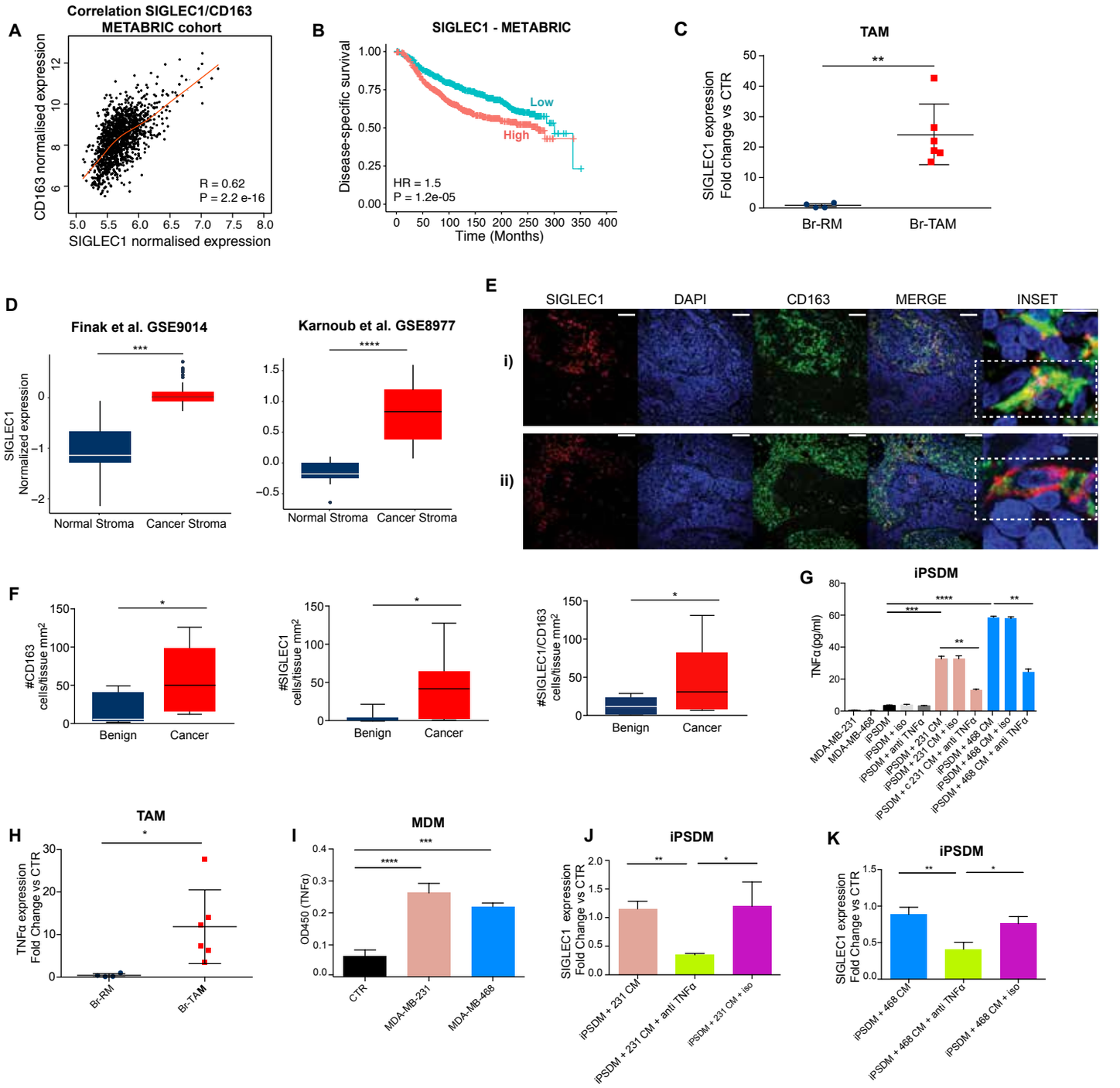

A
Correlation SIGLEC1/CD163
METABRIC cohort
CD163 normalised expression
SIGLEC1 normalised expression
R = 0.62
P = 2.2 e-16
B
SIGLEC1 - METABRIC
Disease-specific survival
Low
High
HR = 1.5
P = 1.2e-05
Time (Months)
C
TAM
SIGLEC1 expression
Fold change vs CTR
Br-RM
Br-TAM
D
Finak et al. GSE9014
Karnoub et al. GSE8977
SIGLEC1
Normalized expression
Normal Stroma
Cancer Stroma
E
SIGLEC1
DAPI
CD163
MERGE
INSET
i)
ii)
F
#CD163
cells/tissue mm2
#SIGLEC1
cells/tissue mm2
#SIGLEC1/CD163
cells/tissue mm2
Benign
Cancer
G
iPSDM
TNFα (pg/ml)
MDA-MB-231
MDA-MB-468
iPSDM
iPSDM + iso
iPSDM + anti TNFα
iPSDM + 231 CM
iPSDM + 231 CM + iso
iPSDM + 231 CM + anti TNFα
iPSDM + c 231 CM + anti TNFα
iPSDM + 468 CM
iPSDM + 468 CM + iso
iPSDM + 468 CM + anti TNFα
H
TAM
TNFα expression
Fold Change vs CTR
Br-RM
Br-TAM
I
MDM
OD450 (TNFα)
CTR
MDA-MB-231
MDA-MB-468
J
iPSDM
SIGLEC1 expression
Fold Change vs CTR
iPSDM + 231 CM
iPSDM + 231 CM + anti TNFα
iPSDM + 231 CM + iso
K
iPSDM
SIGLEC1 expression
Fold Change vs CTR
iPSDM + 468 CM
iPSDM + 468 CM + anti TNFα
iPSDM + 468 CM + iso

Figure 4.

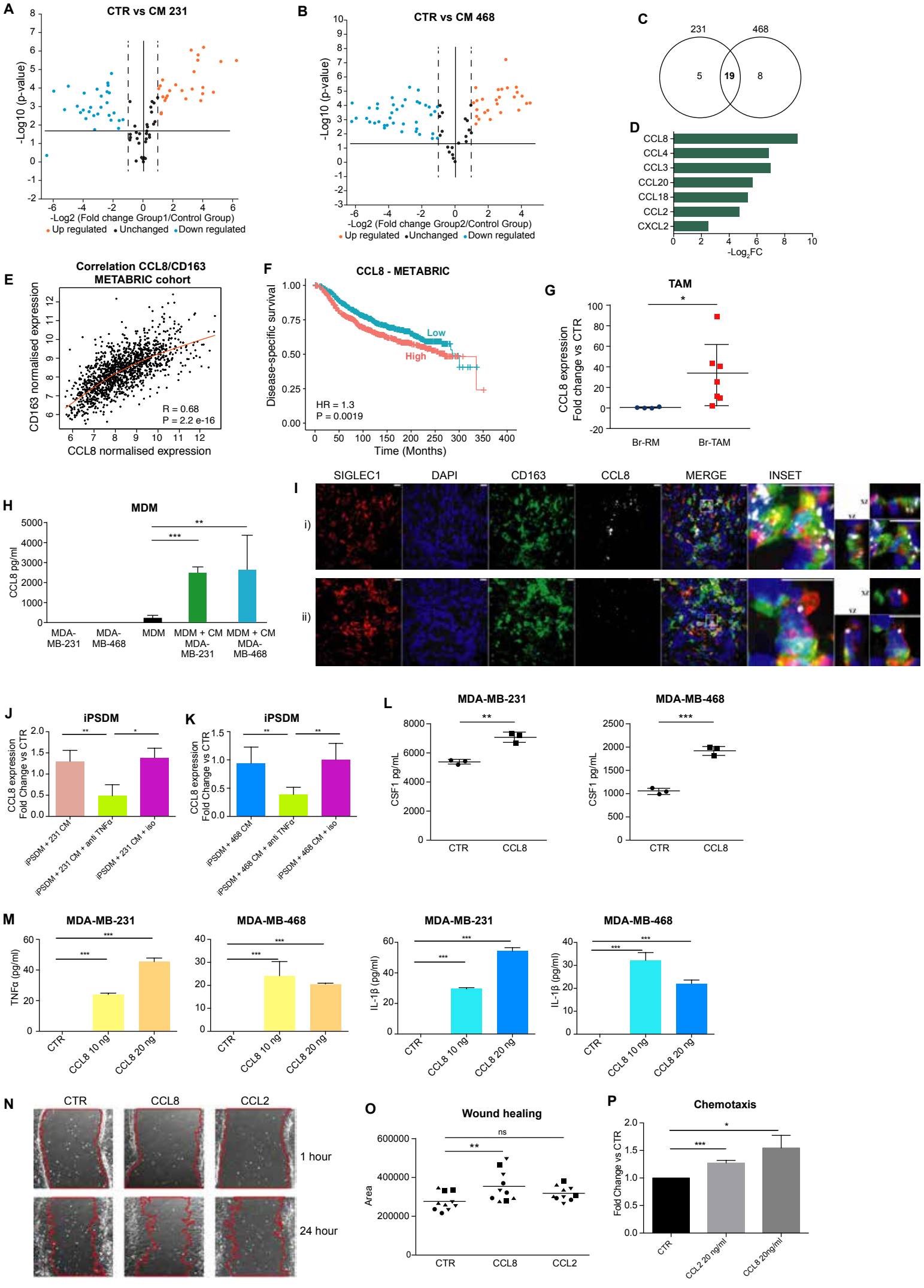

A
CTR vs CM 231
-Log10 (p-value)
-Log2 (Fold change Group1/Control Group)
Up regulated
Unchanged
Down regulated
B
CTR vs CM 468
-Log10 (p-value)
-Log2 (Fold change Group2/Control Group)
Up regulated
Unchanged
Down regulated
C
231
468
5
19
8
D
CCL8
CCL4
CCL3
CCL20
CCL18
CCL2
CXCL2
-Log2FC
E
Correlation CCL8/CD163
METABRIC cohort
CD163 normalised expression
CCL8 normalised expression
R = 0.68
P = 2.2 e-16
F
CCL8 - METABRIC
Disease-specific survival
Low
High
HR = 1.3
P = 0.0019
Time (Months)
G
TAM
CCL8 expression
Fold change vs CTR
Br-RM
Br-TAM
H
MDM
CCL8 pg/ml
MDA-MB-231
MDA-MB-468
MDM
MDM + CM MDA-MB-231
MDM + CM MDA-MB-468
I
SIGLEC1
DAPI
CD163
CCL8
MERGE
INSET
i)
ii)
J
iPSDM
CCL8 expression
Fold Change vs CTR
iPSDM + 231 CM
iPSDM + 231 CM + anti TNFα
iPSDM + 231 CM + iso
K
iPSDM
CCL8 expression
Fold Change vs CTR
iPSDM + 468 CM
iPSDM + 468 CM + anti TNFα
iPSDM + 468 CM + iso
L
MDA-MB-231
CSF1 pg/mL
CTR
CCL8
MDA-MB-468
CSF1 pg/mL
CTR
CCL8
M
MDA-MB-231
TNFα (pg/ml)
CTR
CCL8 10 ng
CCL8 20 ng
MDA-MB-468
TNFα (pg/ml)
CTR
CCL8 10 ng
CCL8 20 ng
MDA-MB-231
IL-1β (pg/ml)
CTR
CCL8 10 ng
CCL8 20 ng
MDA-MB-468
IL-1β (pg/ml)
CTR
CCL8 10 ng
CCL8 20 ng
N
CTR
CCL8
CCL2
1 hour
24 hour
O
Wound healing
Area
CTR
CCL8
CCL2
P
Chemotaxis
Fold Change vs CTR
CTR
CCL2 20 ng/ml
CCL8 20ng/ml

Figure 5.

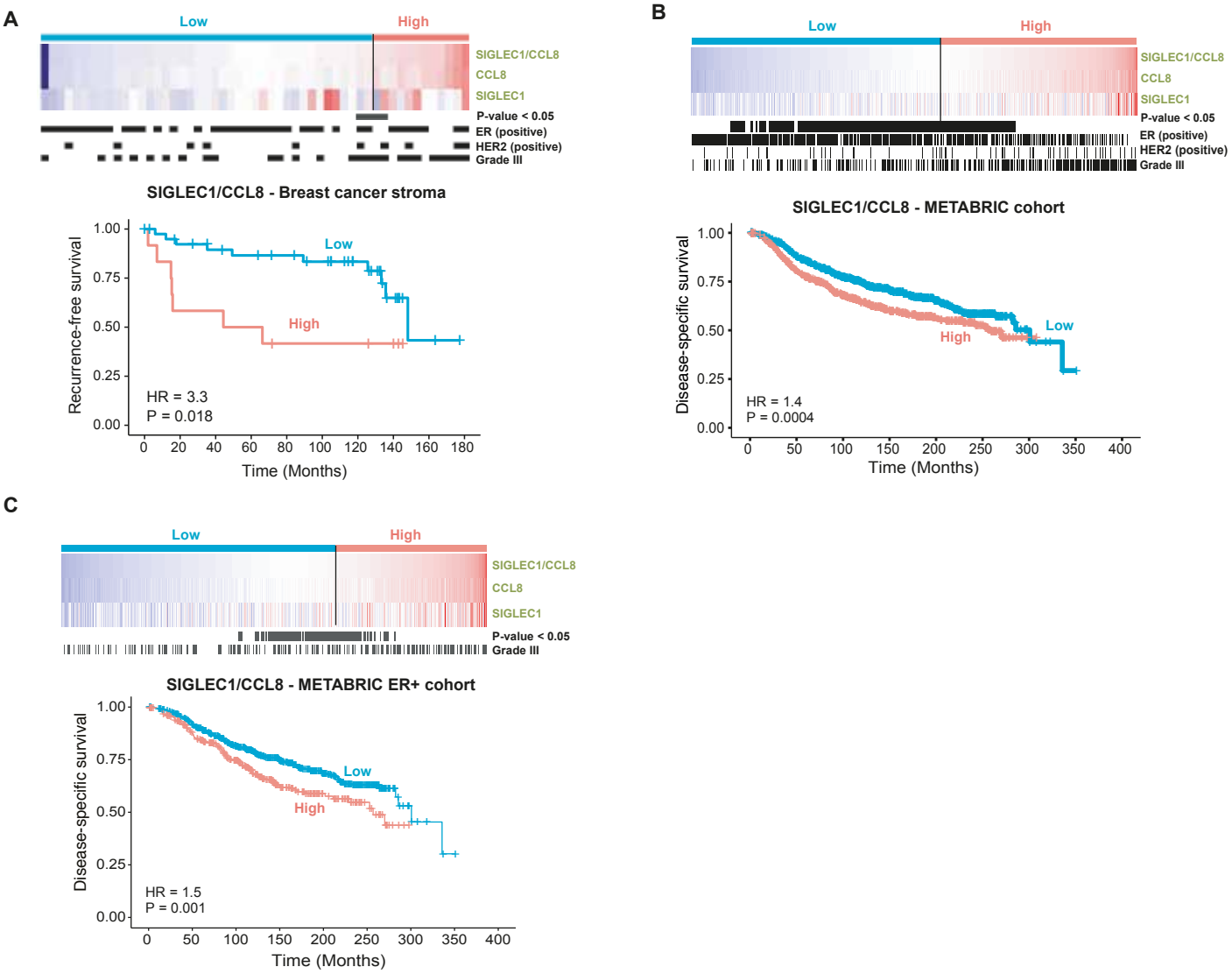

A
Low
High
SIGLEC1/CCL8
CCL8
SIGLEC1
P-value < 0.05
ER (positive)
HER2 (positive)
Grade III
SIGLEC1/CCL8 - Breast cancer stroma
Recurrence-free survival
1.00
0.75
0.50
0.25
0.00
Low
High
HR = 3.3
P = 0.018
0 20 40 60 80 100 120 140 160 180
Time (Months)
B
Low
High
SIGLEC1/CCL8
CCL8
SIGLEC1
P-value < 0.05
ER (positive)
HER2 (positive)
Grade III
SIGLEC1/CCL8 - METABRIC cohort
Disease-specific survival
1.00
0.75
0.50
0.25
0.00
Low
High
HR = 1.4
P = 0.0004
0 50 100 150 200 250 300 350 400
Time (Months)
C
Low
High
SIGLEC1/CCL8
CCL8
SIGLEC1
P-value < 0.05
Grade III
SIGLEC1/CCL8 - METABRIC ER+ cohort
Disease-specific survival
1.00
0.75
0.50
0.25
0.00
Low
High
HR = 1.5
P = 0.001
0 50 100 150 200 250 300 350 400
Time (Months)

Figure 6.

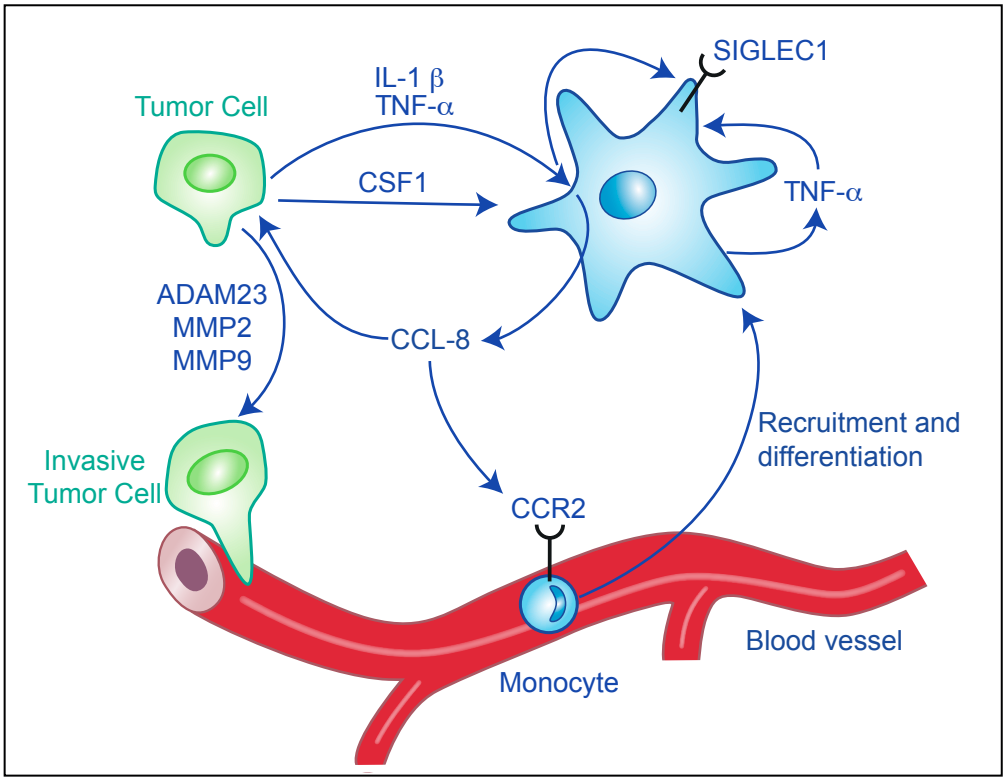

SIGLEC1
IL-1 β
TNF-α
Tumor Cell
CSF1
TNF-α
ADAM23
MMP2
MMP9
CCL-8
Recruitment and
differentiation
Invasive
Tumor Cell
CCR2
Blood vessel
Monocyte

Figure Supp 1.

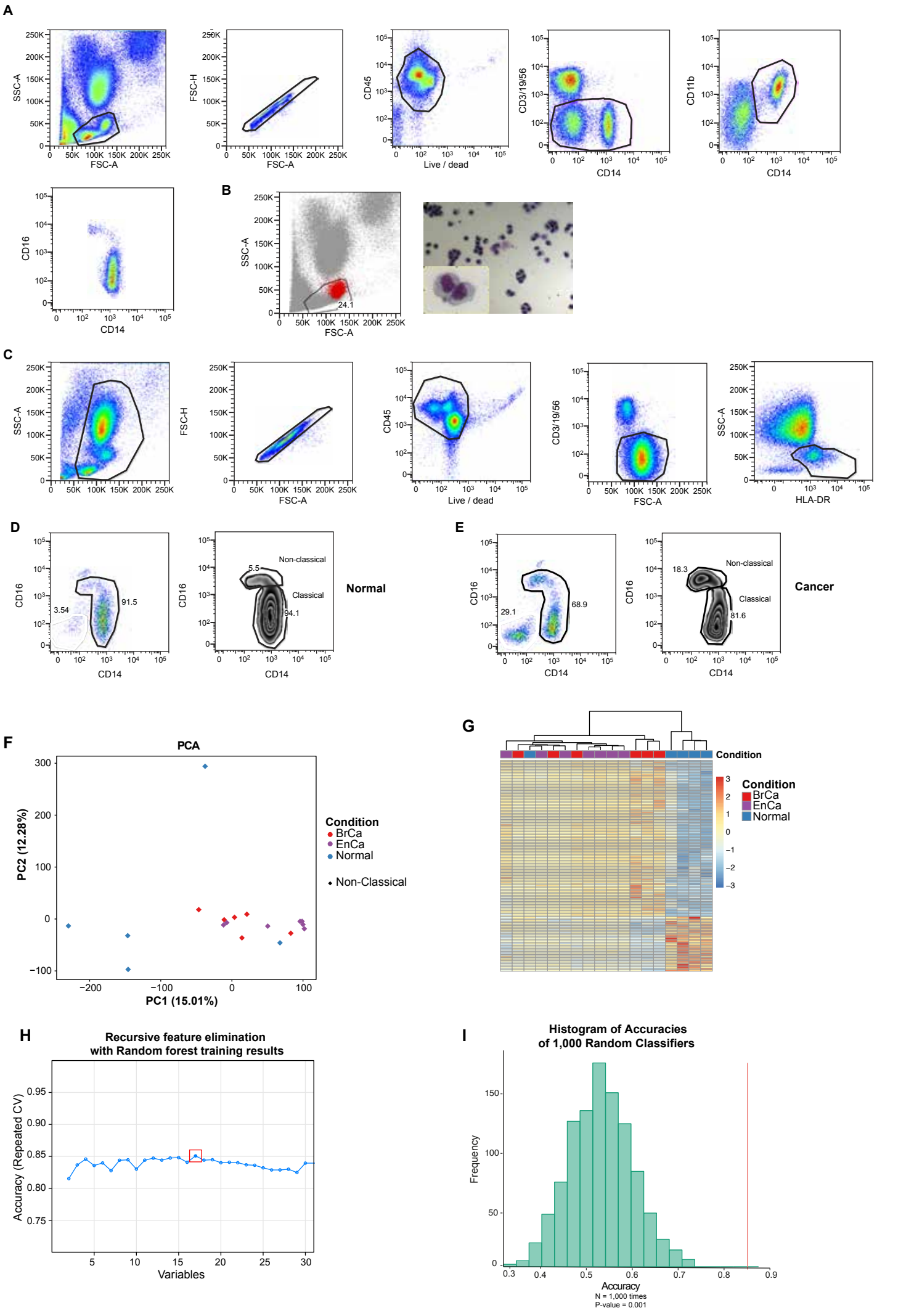


Figure Supp 2.

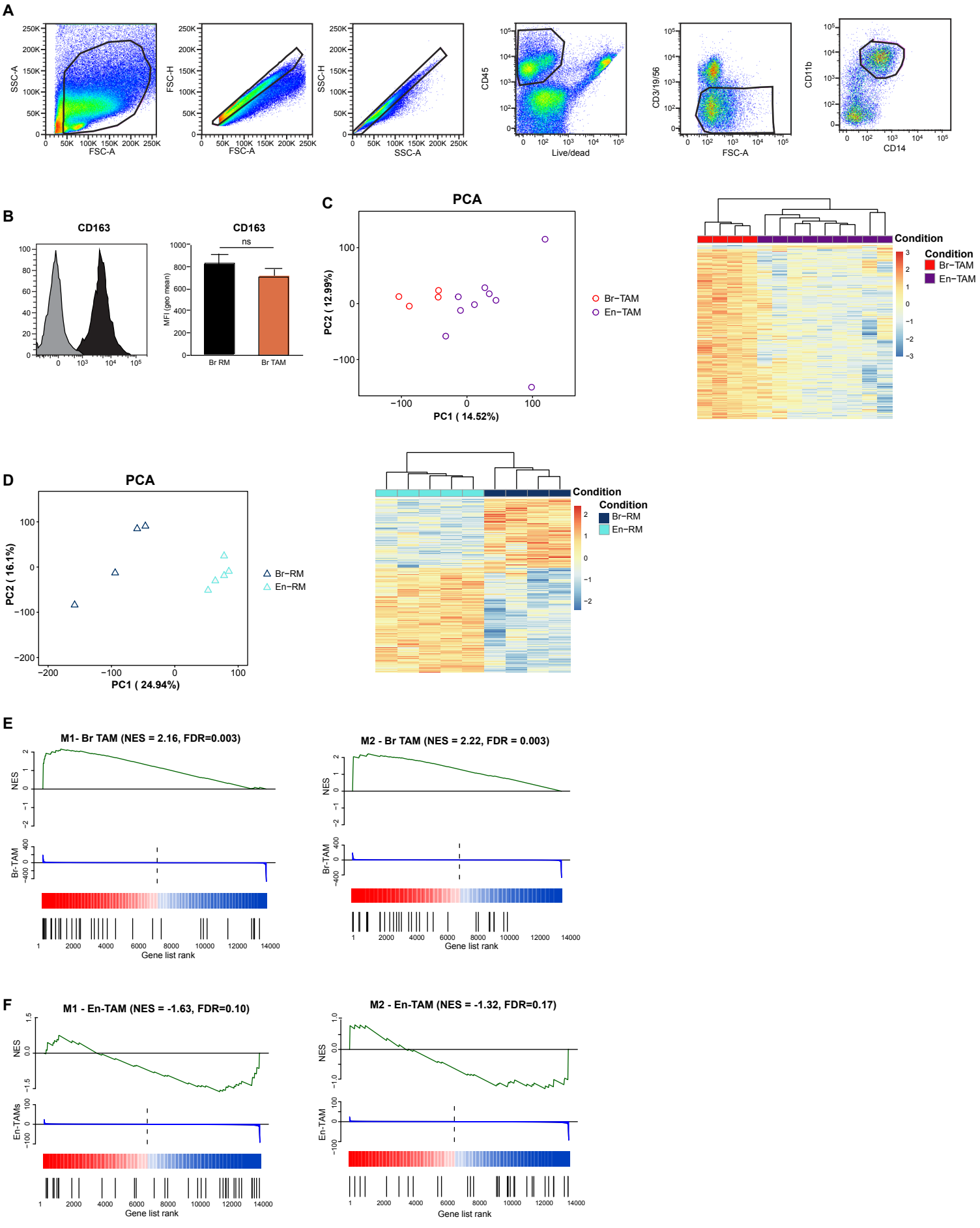
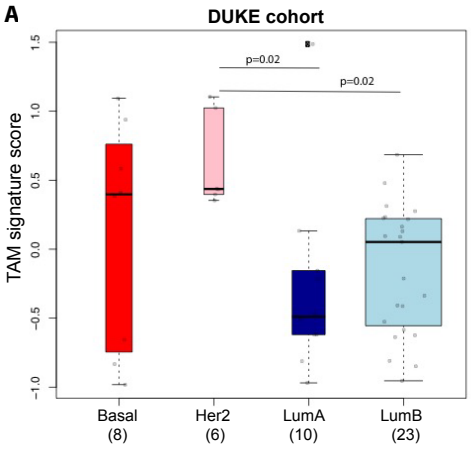
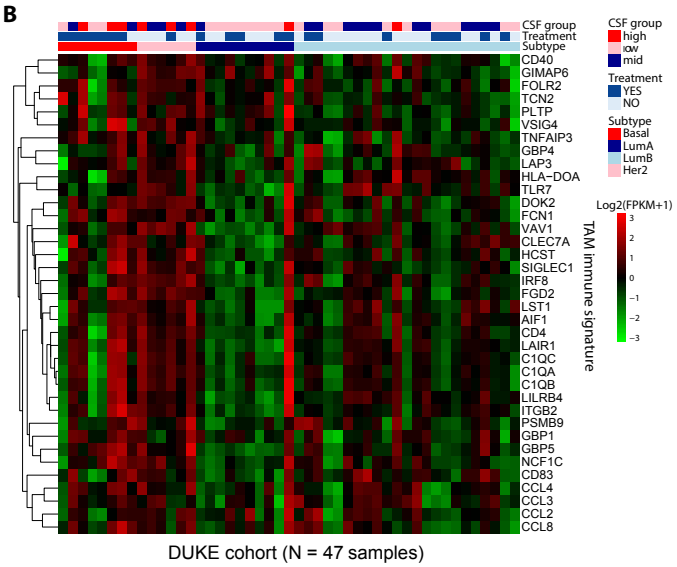
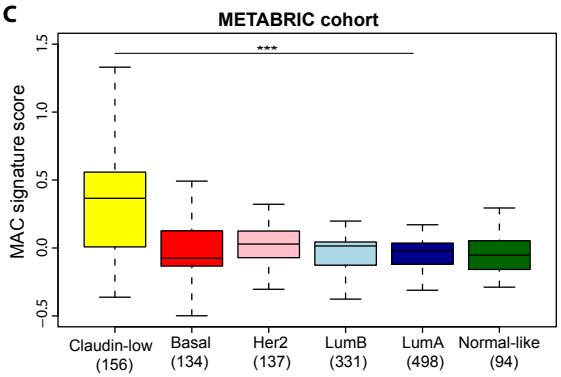


Figure Supp 3.







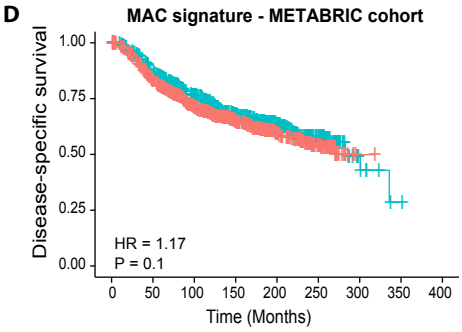


Figure Supp 4.1.

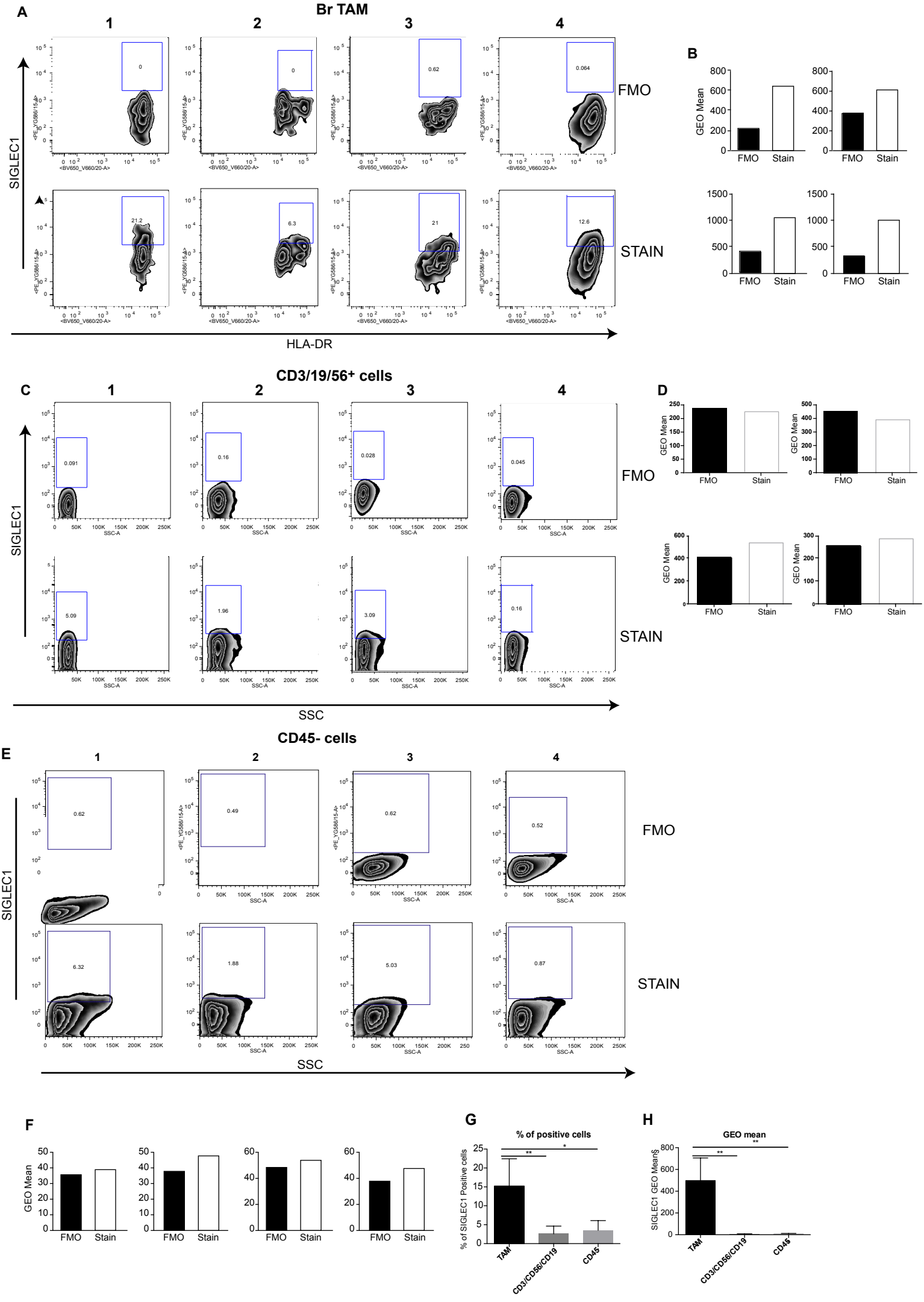


Figure Supp 4.2.

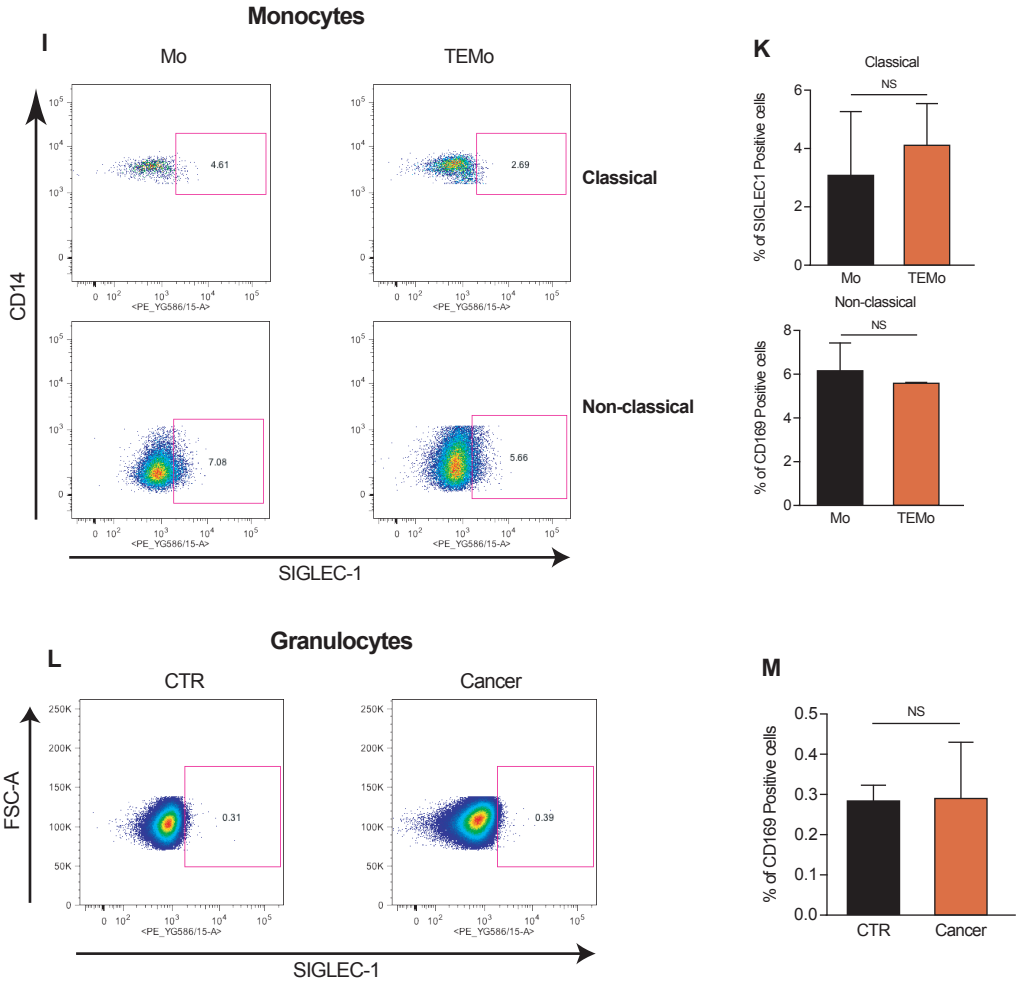
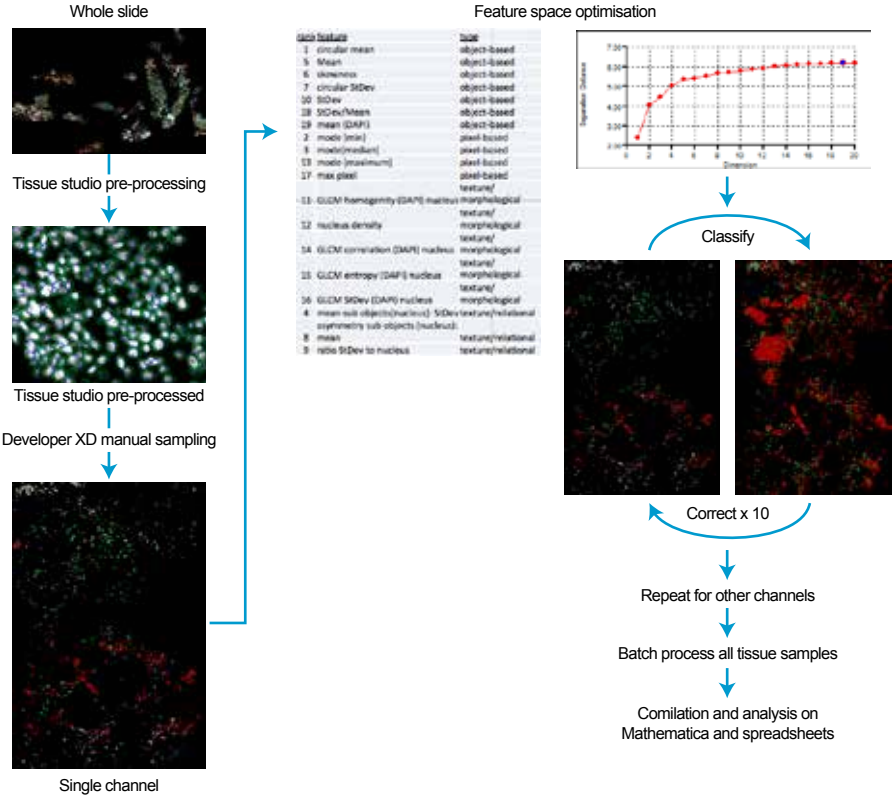
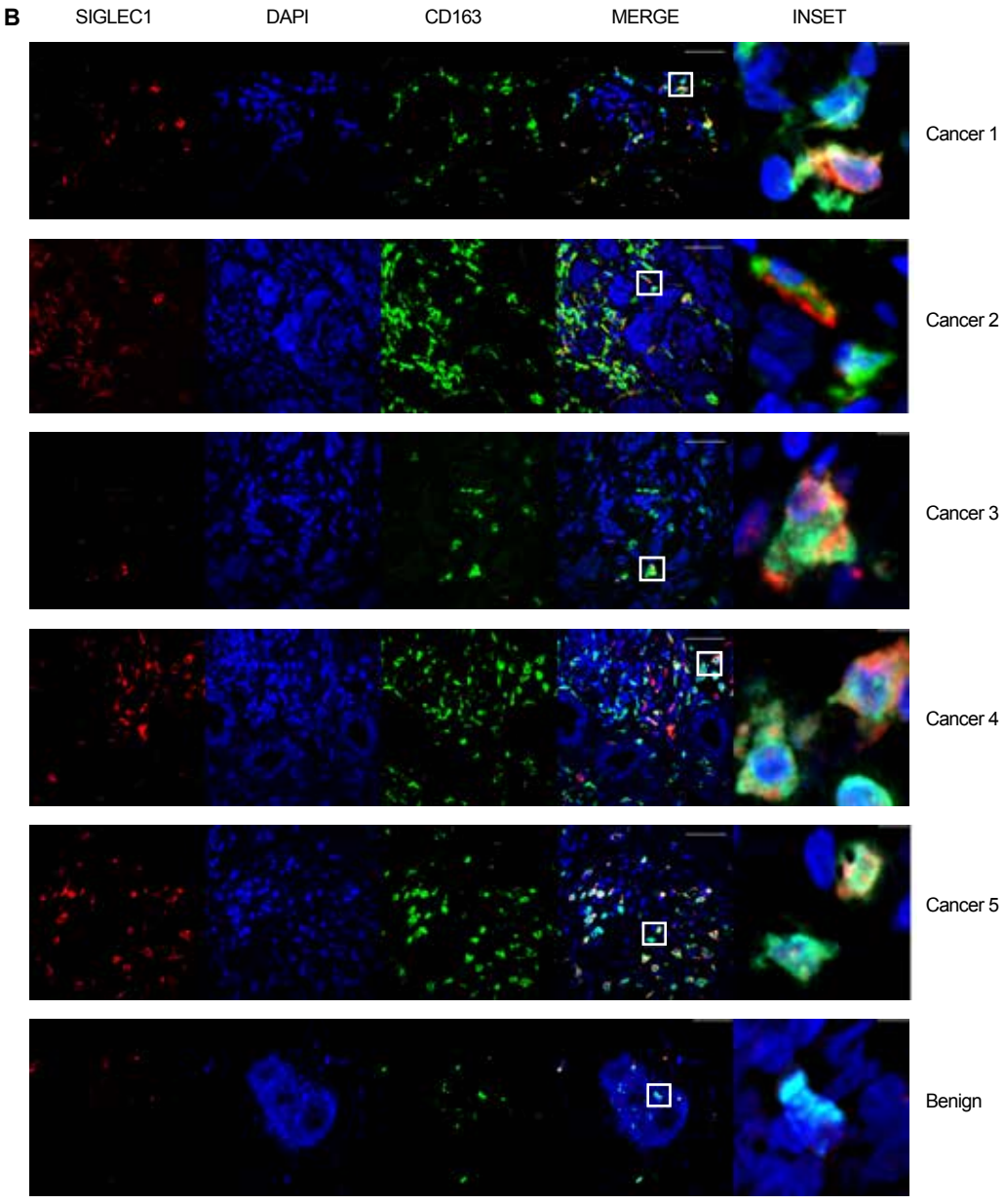


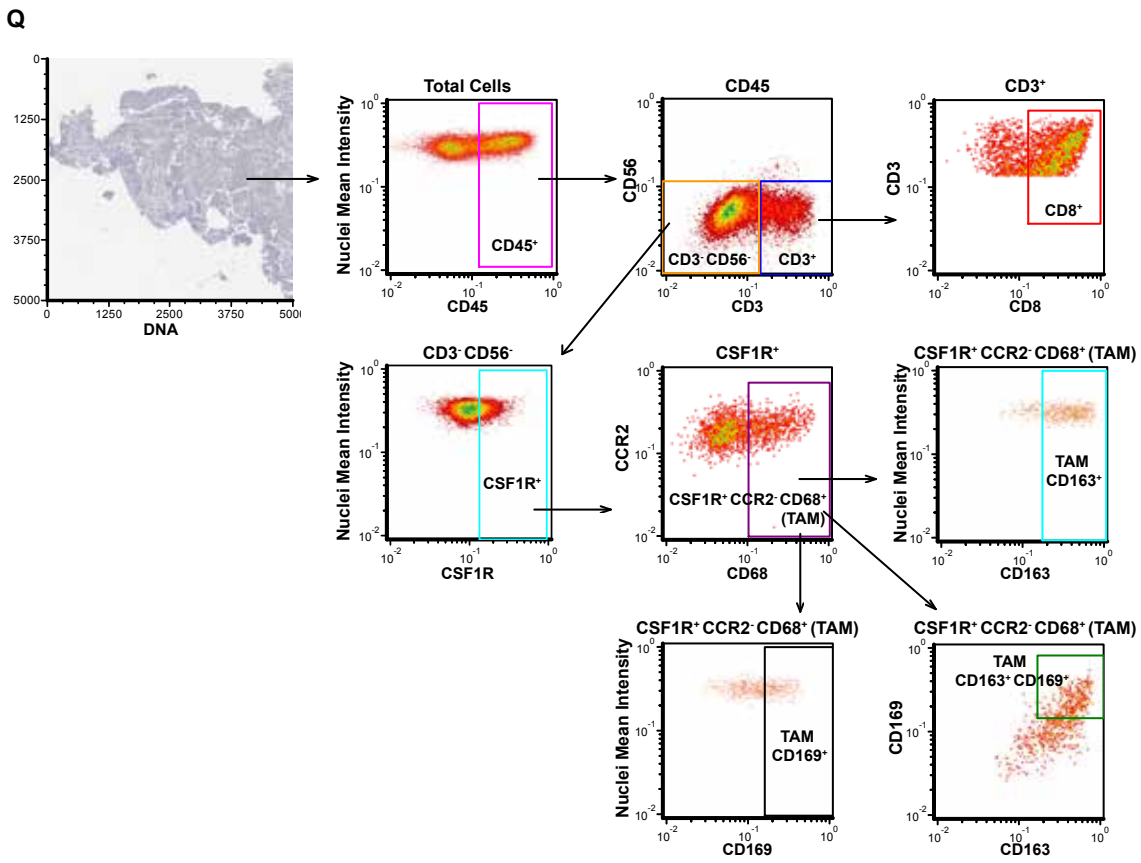
Figure Supp 5.

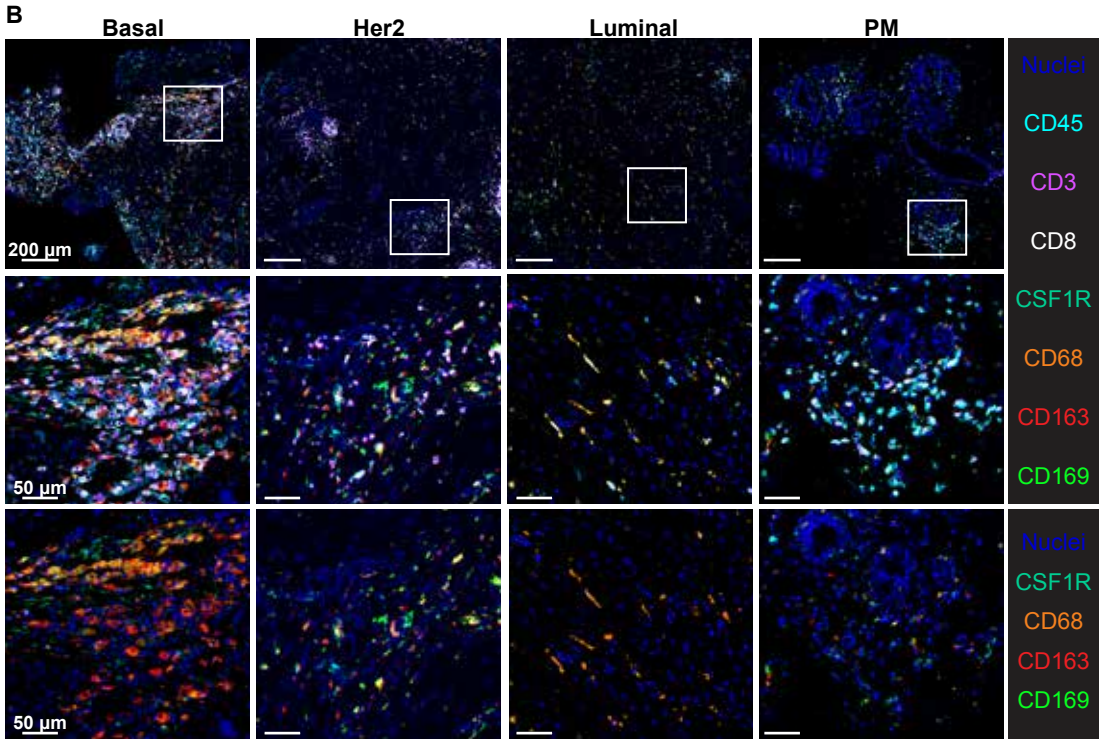
A



B




Q
DNA
Total Cells
Nuclei Mean Intensity
CD45
CD45+
CD56
CD3
CD3- CD56-
CD3+
CD8+
CD8
CSF1R+
CSF1R
CCR2
CD68
CSF1R+ CCR2- CD68+ (TAM)
TAM CD163+
CD163
TAM CD169+
CD169
TAM CD163+ CD169+


B
Basal
Her2
Luminal
PM
Nuclei
CD45
CD3
CD8
CSF1R
CD68
CD163
CD169
200 µm
50 µm
50 µm

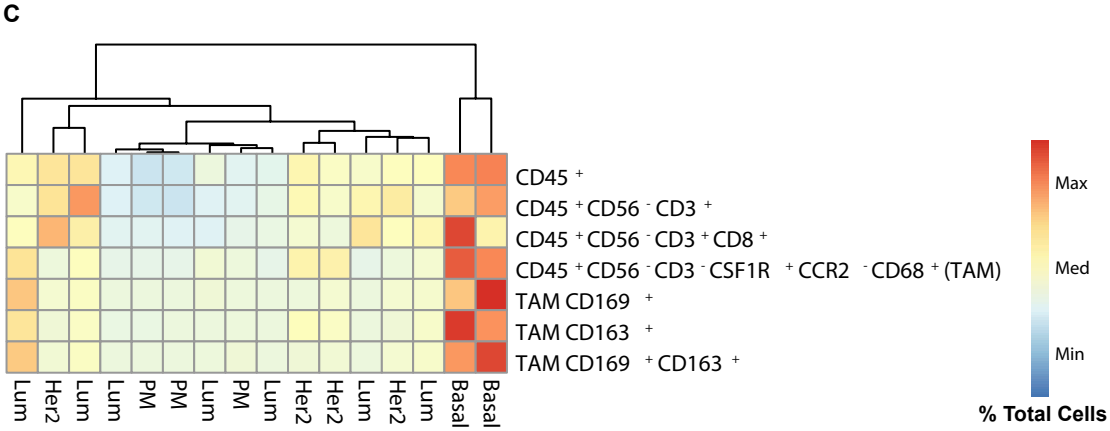

C
CD45 +
CD45 + CD56 - CD3 +
CD45 + CD56 - CD3 + CD8 +
CD45 + CD56 - CD3 - CSF1R + CCR2 - CD68 + (TAM)
TAM CD169 +
TAM CD163 +
TAM CD169 + CD163 +
Lum
Her2
Lum
Lum
PM
PM
Lum
PM
Lum
Her2
Her2
Lum
Her2
Lum
Basal
Basal
Max
Med
Min
% Total Cells

Figure Supp 7.

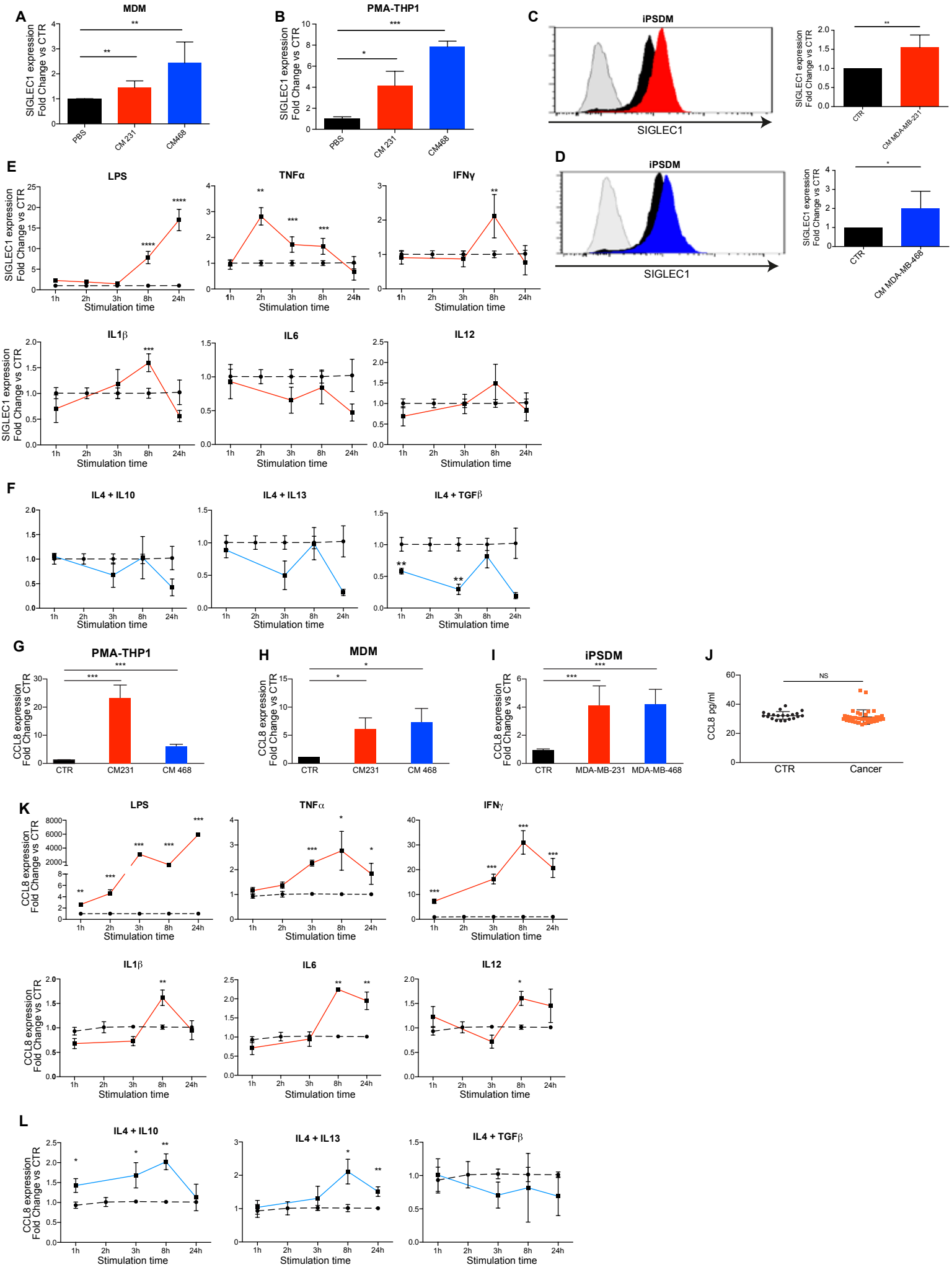


Figure Supp 8.

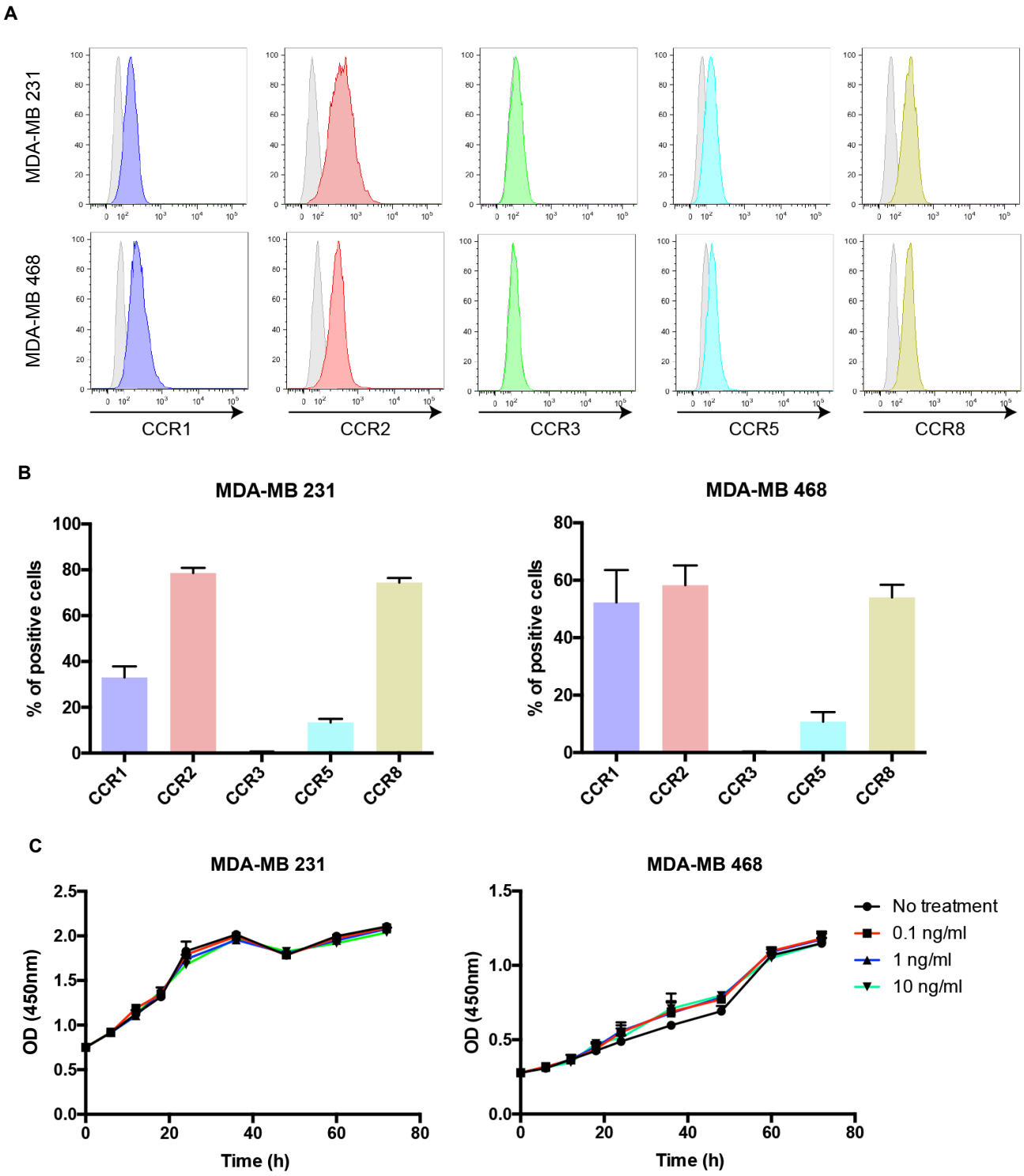


Figure Supp 9.